

# Degree-Degree Dependencies in Random Graphs with Heavy-Tailed Degrees

Remco van der Hofstad and Nelly Litvak

---

**Abstract.** Mixing patterns in large self-organizing networks, such as the Internet, the World Wide Web, social, and biological networks are often characterized by degree-degree dependencies between neighboring nodes. In *assortative* networks, the degree-degree dependencies are positive (nodes with similar degrees tend to connect to each other), whereas in *disassortative* networks, these dependencies are negative. One of the problems with the commonly used Pearson correlation coefficient, also known as the *assortativity coefficient*, is that its magnitude decreases with the network size in disassortative networks. This makes it impossible to compare mixing patterns, for example, in two web crawls of different sizes. As an alternative, we have recently suggested to use rank correlation measures, such as Spearman's rho. Numerical experiments have confirmed that Spearman's rho produces consistent values in graphs of different sizes but similar structure, and it is able to reveal strong (positive or negative) dependencies in large graphs.

In this study we analytically investigate degree-degree dependencies for scale-free graph sequences. In order to demonstrate the ill behavior of the Pearson's correlation coefficient, we first study a simple model of two heavy-tailed, highly correlated,

---

Color versions of one or more of the figures in the article can be found online at [www.tandfonline.com/uinm](http://www.tandfonline.com/uinm).

random variables  $X$  and  $Y$ , and show that the sample correlation coefficient converges in distribution either to a proper random variable on  $[-1, 1]$ , or to zero, and the limit is nonnegative a.s. if  $X, Y \geq 0$ . We next adapt these results to the degree-degree dependencies in networks as described by the Pearson correlation coefficient, and show that it is nonnegative in the large graph limit when the asymptotic degree distribution has an infinite third moment. Furthermore, we provide examples where in the Pearson's correlation coefficient converges to zero in a network with strong negative degree-degree dependencies, and another example where this coefficient converges in distribution to a random variable. We suggest an alternative degree-degree dependency measure, based on Spearman's rho, and prove that this statistical estimator converges to an appropriate limit under quite general conditions. These conditions are proved to be satisfied in common network models, such as the configuration model and the preferential attachment model. We conclude that rank correlations provide a suitable and informative method for uncovering network mixing patterns.

---

## I. Introduction

In this article we present an analytical study of degree-degree correlations in graphs with power-law degree distribution. In simple words, a random variable  $X$  has a power-law distribution with tail exponent  $\gamma > 0$  if its tail probability  $\mathbb{P}(X > x)$  is roughly proportional to  $x^{-\gamma}$ , for large enough  $x$ . Large self-organizing networks, such as the Internet, the World Wide Web, social, and biological networks, usually exhibit high variation in the values of the degrees. Such networks are called *scale free* indicating that there is no typical scale for the degrees, and the high-degree vertices are called *hubs*. This phenomenon is often modeled by using power-law degree distributions.

Power-law distributions are *heavy-tailed* since the tail probability decreases much more slowly than a negative exponential, and, thus, one observes extremely large values of  $X$  much more frequently than in the case of light tails. Statistical analysis of scale-free complex networks has received massive attention in recent literature (see, e.g., [Mitzenmacher 04, Newman 03b] for excellent surveys). Nevertheless, there still are many fundamental open problems. One of them is how to measure *dependencies* between network parameters.

An important characteristic of networks is the dependency between the degrees of direct neighbors. A network is usually called *assortative* when nodes with similar degrees are often connected, thus, the degree-degree dependencies are positive, whereas in a *disassortative* network these dependencies are negative. The *degree-degree* dependencies define many of the network's properties. For instance, the negative degree-degree correlations in the Internet graph have a

great influence on the robustness to failures [Doyle et al. 05], efficiency of Internet protocols [Li et al. 05], as well as distances and betweenness [Mahadevan et al. 06]. The correlation between in- and out-degree of tasks plays an important role in the dynamics of production and development systems [Braha and Bar-Yam 07]. Mixing patterns affect epidemic spread [Eguiluz and Klemm 02, Eubank et al. 04] and Web ranking [Fortunato et al. 07].

Often, degree-degree dependence is characterized by the *assortativity coefficient* of the network, introduced by Newman in [Newman 02]. The assortativity coefficient is in fact the Pearson correlation coefficient between the vector of degrees on each side of an edge, as a function of all edges. See [Newman 02, Table I] for a list of assortativity coefficients for various real-world networks. The empirical data suggest that social networks tend to be assortative (the assortativity coefficient is positive), whereas Internet, World Wide Web, and biological networks tend to be disassortative. In [Newman 02, Table I], it is striking that, typically, larger disassortative networks have assortativity coefficients that are closer to 0 and therefore appear to have approximate *uncorrelated* degrees across edges. Similar conclusions can be drawn from [Newman 03a], see in particular [Newman 03a, Table II]. This phenomenon arises because Pearson's correlation coefficient in scale-free networks with realistic parameters decreases with the network size, as was pointed out in several recent works [Dorogovtsev et al. 10, Raschke et al. 10, van der Hofstad and Litvak 13]. In this study, we prove that Pearson's correlation coefficient in scale-free networks shows several types of pathological behavior; in particular, its infinite volume limit, when it exists, is nonnegative, independently of the mixing pattern, and, in fact, this limit can even be *random*.

In [van der Hofstad and Litvak 13] we propose an alternative measure for the degree-degree dependencies, based on the *ranks* of degrees. This rank correlation approach is in fact classical in multivariate analysis, falling under the category of "concordance measures"—dependency measures based on *order* rather than exact values of two stochastic variables. The huge advantage of such dependency measures is that they work well independently of the number of finite moments of the degrees, whereas Pearson's coefficient suffers from a strong dependence on the extreme values of the degrees. Recent applications of rank correlation measures, such as Spearman's rho [Spearman 04] and the closely related Kendall's tau [Kendall 38], include the concordance between two rankings for a set of documents in web search. In this application field many other measures for rank distances have been proposed, see, for example, [Kumar and Vassilvitskii 10] and the references therein.

We show mathematically that statistical estimators for degree-degree dependencies based on rank correlations are *consistent*. That is, for graphs of different sizes but similar structure (e.g., preferential attachment graphs of increasing size), these estimators converge to their “true” or limiting value that describes the degree-degree dependence in an infinitely large graph (in particular, the variance of the estimator decreases as the size of the graph grows). We also show that Pearson’s correlation coefficient does not have this basic property when degree distributions are heavy-tailed. In particular, as explained in more detail in [van der Hofstad and Litvak 13], this implies that the assortativity coefficient as suggested in [Newman 02] does not allow one to compare the degree-degree dependencies in graphs of different sizes, such as they arise when studying a network at different time stamps, or comparing two different networks, for example, web crawls of different domains or Wikipedia graphs from different languages. However, such a comparison *is* possible using Spearman’s rho. This study forms the mathematical justification of our work [van der Hofstad and Litvak 13], in which similar results were predicted on a less formal level and confirmed by numerical experiments.

This article is organized as follows. In Section 2 we start with the analysis of the sample Pearson correlation coefficient and the sample rank correlation, Spearman’s rho, for a two-dimensional vector with heavy-tailed marginals. In Section 2.3 we present a simple model with an explicit linear dependence and show that the sample size grows to infinity, then Pearson’s correlation coefficient does not converge to a constant but rather to a random variable involving stable distributions. We also verify analytically and numerically that the rank correlation provides a consistent statistical estimator for this model. Next, in Section 2.4, we prove that if random variables are heavy-tailed with infinite second moment and are nonnegative, then the sample Pearson correlation coefficient never converges to a negative value. Thus, such a sequence will never be classified as disassortative. This result is extended to sequences of graphs in Section 3, where we also obtain quite general convergence criteria in the infinite volume limit for the Pearson’s correlation coefficient and the Spearman’s rho. In Section 4 analytical results are provided for Pearson’s correlation coefficient and rank correlations in the configuration model and the preferential attachment model. We also present an adaptation of the configuration model that has strong negative degree-degree dependencies and prove that Spearman’s rho converges to the theoretically justified negative value and Pearson’s coefficient converges to zero. Furthermore, we construct an example in which Pearson’s correlation coefficient converges to a random variable. Numerical results are presented in Section 5. We close the article in Section 6 with a discussion on our results and possible extensions thereof.

## 2. Correlations Between Random Variables

In this section we introduce the dependency measures studied in this work. We start with a general description of dependency measures for random vectors  $(X, Y)$ . This will provide the necessary intuition and framework in order to understand what happens when  $X$  and  $Y$  are the degrees of neighboring nodes in a network graph. We present Pearson’s sample correlation coefficient in Section 2.1, and introduce Spearman’s rho in Section 2.2. In Section 2.3 we demonstrate an ill behavior of Pearson’s sample coefficient in a simple model with linear dependencies, and in Section 2.4 we show that if  $X$  and  $Y$  are nonnegative, then the Pearson’s sample coefficient cannot converge to a negative value.

### 2.1. Sample Pearson’s Correlation Coefficient

The Pearson correlation coefficient  $\rho$  for two random variables  $X$  and  $Y$  with cumulative distribution functions  $F_X(\cdot)$  and  $F_Y(\cdot)$ , joint cumulative distribution function  $F_{X,Y}(\cdot, \cdot)$ , and  $\text{Var}(X), \text{Var}(Y) < \infty$  is defined by

$$\rho = \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}. \tag{2.1}$$

By Cauchy-Schwarz,  $\rho \in [-1, 1]$ , and  $\rho$  measures the *linear dependence* between the random variables  $X$  and  $Y$ . We can approximate  $\rho$  from a sample by computing the *sample correlation coefficient*

$$\rho_n = \frac{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{S_n(X)S_n(Y)}, \tag{2.2}$$

where

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i \tag{2.3}$$

denote the *sample averages* of  $(X_i)_{i=1}^n$  and  $(Y_i)_{i=1}^n$ , while

$$S_n^2(X) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad S_n^2(Y) = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2 \tag{2.4}$$

denote the *sample variances*. For independent identically distributed (i.i.d.) sequences of random vectors  $((X_i, Y_i))_{i=1}^n$  under the assumption of finite-variance random variables, that is,  $\text{Var}(X), \text{Var}(Y) < \infty$ , it is well known that the estimator  $\rho_n$  of  $\rho$  is consistent, in other words,

$$\rho_n \xrightarrow{\mathbb{P}} \rho, \tag{2.5}$$

where  $\xrightarrow{\mathbb{P}}$  denotes convergence in probability. In practice, however, we tend not to know whether  $\text{Var}(X), \text{Var}(Y) < \infty$ , since  $S_n^2(X) < \infty$  and  $S_n^2(Y) < \infty$  clearly hold for any sample, and, therefore, one might be tempted to always use  $\rho_n$ . Furthermore, by the Cauchy-Schwarz inequality,  $\rho_n \in [-1, 1]$  for every  $n \geq 1$ , which is part of the problem, because, for any sample, a value in  $[-1, 1]$  is produced, and no alarm bells start ringing when  $\rho_n$  is used inappropriately. In this work we investigate the case  $\text{Var}(X), \text{Var}(Y) = \infty$  and show that the use of  $\rho_n$ , in this case, and in particular in scale-free random graphs, is uninformative. For example, in the case of negative correlations,  $\rho_n$  converges to zero when  $n \rightarrow \infty$ , which makes it impossible to compare the data of different sizes. Moreover, if correlations are positive,  $\rho_n$  may even converge to a random variable, thus it can produce very different numbers for two random structures of the same size created by the same mechanism. We provide such examples for linearly dependent random variables in Section 2.3 and for random graphs in Section 4.4.

## 2.2. Rank Correlations

For two-dimensional data  $((X_i, Y_i))_{i=1}^n$ , let  $r_i^X$  and  $r_i^Y$  be the rank of an observation  $X_i$  and  $Y_i$ , respectively, when the sample values  $(X_i)_{i=1}^n$  and  $(Y_i)_{i=1}^n$  are arranged in a descending order. The idea of rank correlations is in evaluating statistical dependences on the data  $((r_i^X, r_i^Y))_{i=1}^n$ , rather than on the original data  $((X_i, Y_i))_{i=1}^n$ . Rank transformation is convenient, in particular because, for continuous random variables, the two marginals of the resulting vector  $(r_i^X, r_i^Y)$  are realizations of identical uniform distributions, implying many nice mathematical properties.

The statistical correlation coefficient for the ranks is known as Spearman's rho [Spearman 04]:

$$\begin{aligned} \rho_n^{\text{rank}} &= \frac{\sum_{i=1}^n (r_i^X - (n+1)/2)(r_i^Y - (n+1)/2)}{\sqrt{\sum_{i=1}^n (r_i^X - (n+1)/2)^2 \sum_{i=1}^n (r_i^Y - (n+1)/2)^2}} \\ &= \frac{\frac{1}{n} \sum_{i=1}^n r_i^X r_i^Y - ((n+1)/2)^2}{\frac{1}{12} (n^2 - 1)}. \end{aligned} \quad (2.6)$$

The mathematical properties of Spearman's rho have been extensively investigated in the literature. It is well known that if  $((X_i, Y_i))_{i=1}^n$  consists of independent realizations of  $(X, Y)$ , and the joint distribution cumulative function of  $X$  and  $Y$  is continuous, then  $\rho_n^{\text{rank}}$  converges to a number that can be interpreted as its population value, see [Kendall 75, Chapter 9, Borkowf 02]:

$$\rho_n^{\text{rank}} \xrightarrow{\mathbb{P}} \rho^{\text{rank}} = 12\mathbb{E}(F_X(X)F_Y(Y)) - 3. \quad (2.7)$$

For completeness, we give a brief explanation of this formula. Observe that  $F_X(X)$  is the random variable that takes the value  $F_X(x)$  when  $X = x$ . If  $X$

is continuous, then  $F_X(X)$  has a uniform distribution on  $[0, 1]$ :

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(F_X(X) \leq F_X(x)). \tag{2.8}$$

Now take  $F_X(x) = t$  to obtain  $\mathbb{P}(F_X(X) \leq t) = t$ , where  $t$  can take any value in  $[0, 1]$ . We note that this derivation holds for *any* continuous random variable  $X$ . We will use this many times throughout the article. In particular, it follows that  $\mathbb{E}(F_X(X)) = \mathbb{E}(F_Y(Y)) = 1/2$ . Next, note that  $r_i^X/n$  is an empirical estimator of  $1 - F_X(x_i)$ , where  $x_i$  is the realized value of  $X_i$ . Moreover,

$$\begin{aligned} \mathbb{E}((1 - F_X(X))(1 - F_Y(Y))) &= 1 - \mathbb{E}(F_X(X)) - \mathbb{E}(F_Y(Y)) + \mathbb{E}(F_X(X)F_Y(Y)) \\ &= \mathbb{E}(F_X(X)F_Y(Y)). \end{aligned}$$

Hence, the right-hand side of (2.6) is a statistical estimator of the last expression in (2.7).

For discrete random variables, the situation is more delicate, because the same values of  $X$  and  $Y$  could occur more than once. We resolve the ties randomly, using uniformization as suggested in [Mesfioui and Tajar 05]. Formally, we replace the ranks of  $((X_i, Y_i))_{i=1}^n$  by the ranks of the random variables

$$((X_i^*, Y_i^*))_{i=1}^n = ((X_i + U_i, Y_i + U'_i))_{i=1}^n,$$

where  $((U_i, U'_i))_{i=1}^n$  is a sequence of  $2n$  i.i.d. uniform variables on  $(0, 1)$ . The random variables  $X_i^*$  and  $Y_i^*$  now *are* continuous. We denote their cumulative distribution functions by  $F_X^*$  and  $F_Y^*$ . Note that if  $X$  takes nonnegative integer values, then  $F_X^*$  can be seen as a linear interpolation of the cumulative probability  $\mathbb{P}(X < x)$ ,  $x = 0, 1, 2, \dots$  because  $\mathbb{P}(X = x) = \mathbb{P}(X^* \in [x, x + 1))$ .

Since  $(X^*, Y^*)$  has a continuous distribution, the convergence result in (2.7) remains valid. Moreover, [Mesfioui and Tajar 05] gives the formula for  $\rho^{\text{rank}}$  in a discrete case, and [Mesfioui and Tajar 05, Proposition 3.1] states that if  $X, Y = 0, 1, \dots$ , then  $(X, Y)$  and  $(X^*, Y^*)$  have the population value  $\rho^{\text{rank}}$ , i.e.,

$$\rho^{\text{rank}} = 12\mathbb{E}(F_X^*(X^*)F_Y^*(Y^*)) - 3. \tag{2.9}$$

The comparison of different ways for resolving ties, and their effect on the resulting computation is an interesting topic, which is outside the scope of this work. We refer the reader to [Nevslehová 07] for a general treatment of rank correlations for noncontinuous distributions.

### 2.3. Linear Dependencies

It is well known that  $\rho$  in general measures *linear* dependence between two random variables. Therefore, before analyzing the behavior of  $\rho_n$  in networks, we wish to illustrate that  $\rho_n$  fails to capture the linear dependence between  $X$  and  $Y$  when the variances of  $X$  and  $Y$  are infinite, that is,  $\text{Var}(X), \text{Var}(Y) = \infty$ ,

even in a very straightforward case when the linear relation between  $X$  and  $Y$  is explicitly defined. With this goal in mind, we analyze the behavior of  $\rho_n$  in the following linear model:

$$X = \alpha_1 \xi_1 + \cdots + \alpha_m \xi_m, \quad Y = \beta_1 \xi_1 + \cdots + \beta_m \xi_m, \quad (2.10)$$

where  $\xi_j, j = 1, \dots, m$ , are i.i.d. nonnegative random variables with regularly varying tail, and tail exponent  $\gamma$ . By definition, the nonnegative random variable  $\xi$  is *regularly varying* with index  $\gamma > 0$ , if

$$\mathbb{P}(\xi > x) = L(x)x^{-\gamma}, \quad x \geq 0, \quad (2.11)$$

where  $x \mapsto L(x)$  is a slowly varying function, that is, for  $u > 0, L(ux)/L(x) \rightarrow 1$  as  $x \rightarrow \infty$ , for instance,  $L(x)$  may be equal to a constant or  $\log(x)$ . Note that the random variables  $X$  and  $Y$  have the same distribution when  $(\beta_1, \dots, \beta_m)$  is a permutation of  $(\alpha_1, \dots, \alpha_m)$ .

When we take an i.i.d. sample of random variables  $((X_i, Y_i))_{i=1}^n$  with the above linear dependence, then Spearman's rho is consistent by (2.7), with a variance that converges to zero as  $1/n$ . For the sample correlation coefficient, consistency follows from (2.5) in, the case where  $\text{Var}(\xi_i) < \infty$ , but not when the  $\xi_i$ 's have infinite variance as we show in detail following. Our main result in this section is the following theorem:

**Theorem 2.1. (Weak convergence of the sample Pearson's coefficient).** *Let  $((X_i, Y_i))_{i=1}^n$  be i.i.d. copies of the random variables  $(X, Y)$  in (2.10), and where  $(\xi_j)_{j=1}^m$  are i.i.d. random variables satisfying (2.11) with  $\gamma \in (0, 2)$ , so that  $\text{Var}(\xi_j) = \infty$ . Then,*

$$\rho_n \xrightarrow{d} \rho \equiv \frac{\sum_{j=1}^m \alpha_j \beta_j Z_j}{\sqrt{\sum_{j=1}^m \alpha_j^2 Z_j} \sqrt{\sum_{j=1}^m \beta_j^2 Z_j}}, \quad (2.12)$$

where  $(Z_j)_{j=1}^m$  are i.i.d. random variables having stable distributions with parameter  $\gamma/2 \in (0, 1)$ , and  $\xrightarrow{d}$  denotes convergence in distribution. In particular,  $\rho$  has a density on  $[-1, 1]$ . This density is strictly positive on  $(-1, 1)$  when there exist  $k, l$  such that  $\alpha_k \beta_k < 0 < \alpha_l \beta_l$ . Furthermore, the density is positive on  $(a, 1)$  when  $\alpha_k \beta_k \geq 0$  for every  $k$ , and on  $(-1, -a)$  when  $\alpha_k \beta_k \leq 0$  for every  $k$ , where

$$a = \inf_{z_1, \dots, z_m \in \mathbb{R}} \frac{\sum_{j=1}^m |\alpha_j \beta_j| z_j}{\sqrt{\sum_{j=1}^m \alpha_j^2 z_j} \sqrt{\sum_{j=1}^m \beta_j^2 z_j}} \in (0, 1). \quad (2.13)$$

Theorem 2.1 states that the sample correlation coefficient converges in distribution to a proper random variable, contrary to Spearman's rank correlation, which converges in probability to a constant. In particular, this implies that



when we have two independent samples, the sample correlation coefficient will give two rather distinct values, whereas Spearman’s rank correlation will give two similar values. We prove Theorem 2.1 in the remainder of this section. In its proof, we need the following technical result:

**Lemma 2.2. (Asymptotics of sums in stable domain).** *Let  $(\xi_{i,j})_{i=1,2,\dots,n,j=1,2}$  be i.i.d. random variables satisfying (2.11) for some  $\gamma \in (0, 2)$ . Then there exists a sequence  $a_n$  with  $a_n = n^{2/\gamma} \ell(n)$ , where  $n \mapsto \ell(n)$  is slowly varying, such that*

$$\frac{1}{a_n} \sum_{i=1}^n \xi_{i,1}^2 \xrightarrow{d} Z_1, \quad \frac{1}{a_n} \sum_{i=1}^n \xi_{i,1} \xi_{i,2} \xrightarrow{\mathbb{P}} 0, \tag{2.14}$$

where  $Z_1$  is stable with parameter  $\gamma/2$  and  $\xrightarrow{\mathbb{P}}$  denotes convergence in probability.

**Proof.** Let  $F(x) = \mathbb{P}(\xi \leq x)$  be the cumulative distribution function of  $\xi$ . In order to prove the first statement in (2.14) we need to note only that the cumulative distribution function of  $\xi^2$  equals  $x \mapsto F(\sqrt{x})$ , which, by (2.11), implies that  $\xi^2$  is regularly varying. Thus, the first statement in (2.14) is in fact the classical convergence of infinite variance random variables with slowly varying distribution functions to stable laws (see e.g., [Gnedenko and Kolmogorov 68]), where  $Z_1$  is a stable  $\gamma/2$  random variable. In particular, denoting  $[1 - F](x) = 1 - F(x)$ ,  $x \geq 0$ , we can identify  $a_n = [1 - F]^{-1}(1/n^2)$  [Bingham et al. 89]. Since  $x \mapsto [1 - F](x)$  is regularly varying with index  $\gamma$ ,  $[1 - F]^{-1}(1/n)$  is regularly varying with index  $1/\gamma$  [Bingham et al. 89], so that  $a_n = [1 - F]^{-1}(1/n^2)$  is regularly varying with index  $2/\gamma$ . To prove the second part of (2.14), we write

$$1 - F(x) = \mathbb{P}(\xi > x) \leq c' x^{-\gamma'}, \quad x \geq 0, \tag{2.15}$$

which is valid for any  $\gamma' \in (1, \gamma)$  by (2.11) and Potter’s theorem. We next study the cumulative distribution function of  $\xi_1 \xi_2$ , which we denote by  $H$ , where  $\xi_1$  and  $\xi_2$  are two independent copies of the random variable  $\xi$ . When  $F$  satisfies (2.15), then it is not hard to see that there exists a  $C > 0$  such that

$$1 - H(u) \leq C(1 + \log u)u^{-\gamma'}. \tag{2.16}$$

Indeed, assume that  $F$  has a density  $f(w) = cw^{-(\gamma'+1)}$ , for  $w \geq 1$ . Then,

$$1 - H(u) = \int_1^\infty f(w)[1 - F](u/w)dw.$$

Clearly,  $1 - F(w) = c'w^{-\gamma'}$  for  $w \geq 1$  and  $1 - F(w) = 1$  otherwise. Substitution of this yields

$$1 - H(u) \leq cc' \int_1^u w^{-(\gamma'+1)}(u/w)^{-\gamma'} dw + c \int_u^\infty w^{-(\gamma'+1)} dw \leq C(1 + \log u)u^{-\gamma'}.$$

When  $F$  satisfies (2.15), then  $\xi_1$  and  $\xi_2$  are stochastically upper bounded by  $\hat{\xi}_1$  and  $\hat{\xi}_2$  with cumulative distribution function  $\hat{F}$  satisfying  $1 - \hat{F}(w) = c'w^{-\gamma'} \vee 1$ , where  $(x \vee y) = \max\{x, y\}$ , and the claim in (2.16) follows from the above computation.

By the bound in (2.16), the random variables  $\xi_{i,1}\xi_{i,2}$  are stochastically bounded from above by random variables  $P_i$  that are in the domain of attraction of a stable  $\gamma'$  random variable. As a result, there exists  $b_n = n^{1/\gamma'}\ell'(n)$ , where  $n \mapsto \ell'(n)$  is slowly varying, such that

$$\frac{1}{b_n} \sum_{i=1}^n P_i \xrightarrow{d} W,$$

where  $W$  is stable  $\gamma'$ . By choosing  $\gamma' > \gamma/2$ , we get  $b_n/a_n \rightarrow 0$ , so we obtain the second statement in (2.14). □

**Proof of Theorem 2.1.** We start by noting that

$$\rho_n = \frac{\frac{1}{n-1} \sum_{i=1}^n (X_i Y_i - \bar{X}_n \bar{Y}_n)}{S_n(X) S_n(Y)}, \tag{2.17}$$

and

$$S_n^2(X) = \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - \bar{X}_n^2), \quad S_n^2(Y) = \frac{1}{n-1} \sum_{i=1}^n (Y_i^2 - \bar{Y}_n^2). \tag{2.18}$$

We continue to identify the asymptotic behavior of

$$\sum_{i=1}^n X_i^2, \quad \sum_{i=1}^n Y_i^2, \quad \sum_{i=1}^n X_i Y_i.$$

Let  $[n]$  denote the set of integers  $\{1, 2, \dots, n\}$ . The distribution of  $((X_i, Y_i))_{i=1}^n$  is described in terms of an array  $(\xi_{i,j})_{i \in [n], j \in [m]}$ , which are i.i.d. copies of a random variable  $\xi$ . In terms of these random variables, we can identify

$$\sum_{i=1}^n X_i Y_i = \sum_{j=1}^m \alpha_j \beta_j \left( \sum_{i=1}^n \xi_{i,j}^2 \right) + \sum_{j_1 \neq j_2=1}^m \alpha_{j_1} \beta_{j_2} \left( \sum_{i=1}^n \xi_{i,j_1} \xi_{i,j_2} \right). \tag{2.19}$$

The sums  $\sum_{i=1}^n \xi_{i,j}^2$  are i.i.d. for different  $j \in \{1, \dots, m\}$ , and by Lemma 2.2,  $\sum_{i=1}^n \xi_{i,j_1} \xi_{i,j_2}$  is of a smaller order. Hence, from (2.19) we obtain that

$$\frac{1}{a_n} \sum_{i=1}^n X_i Y_i \xrightarrow{d} \sum_{j=1}^m \alpha_j \beta_j Z_j. \tag{2.20}$$

Therefore, by taking  $\alpha = \beta$ , we also obtain

$$\frac{1}{a_n} \sum_{i=1}^n X_i^2 \xrightarrow{d} \sum_{j=1}^m \alpha_j^2 Z_j, \quad \frac{1}{a_n} \sum_{i=1}^n Y_i^2 \xrightarrow{d} \sum_{j=1}^m \beta_j^2 Z_j, \quad (2.21)$$

and the convergence holds *simultaneously*. As a result, (2.12) follows. It remains to establish the properties of the limiting random variable  $\rho$  in (2.12).

The density of  $Z_i$  is strictly positive on  $(0, \infty)$ . Note that rescaling  $z_j = cz_j$   $j = 1, \dots, m$ , in (2.13), does not change the value of  $a$ . In particular, we can choose  $c = (\max\{z_1, z_2, \dots, z_m\})^{-1}$ . If there exist  $k$  and  $l$  such that  $\alpha_k \beta_k < 0 < \alpha_l \beta_l$ , then the density of  $\rho$  is strictly positive on  $(-1, 1)$ . Indeed, with positive probability  $\rho$  can be arbitrarily close to  $-1$  if  $Z_k = \max\{Z_1, \dots, Z_m\}$  and  $Z_j/Z_k$ ,  $j \neq k$  are sufficiently small. Similarly, if  $Z_l = \max\{Z_1, \dots, Z_m\}$ , then with positive probability,  $\rho$  can be arbitrarily close to  $1$ . Now assume that  $\alpha_k \beta_k \geq 0$  for every  $k$ . In this case, the density of  $\rho$  is strictly positive on the support of  $\rho$ , which is  $(a, 1)$ , with  $a$  as in (2.13). Analogously, when  $\alpha_k \beta_k \leq 0$ , then  $\rho$  cannot be positive and has a density on  $(-1, -a)$ .  $\square$

**Numerical example.** In order to illustrate the result of Theorem 2.1, consider the example with  $\xi_j$ 's from a Pareto distribution satisfying  $\mathbb{P}(\xi > x) = 1/x^{1.1}$ ,  $x \geq 1$ , so  $L(x) = 1$  and  $\gamma = 1.1$  in (2.11). The exponent  $\gamma = 1.1$  is as observed for the World Wide Web [Broder et al. 00]. In (2.10), we choose  $m = 3$  and  $\alpha_i, \beta_i, i = 1, 2, 3$ , as specified in Table 1. We generate  $N$  data samples  $((X_i, Y_i))_{i=1}^n$  and compute  $\rho_n$  and  $\rho_n^{\text{rank}}$  for each of the  $N$  samples. Thus, we obtain the vectors  $(\rho_{n,j})_{j=1}^N$  and  $(\rho_{n,j}^{\text{rank}})_{j=1}^N$  of  $N$  independent realizations for  $\rho_n$  and  $\rho_n^{\text{rank}}$ , respectively, where the subindex  $j = 1, \dots, N$  denotes the  $j$ th realization of  $((X_i, Y_i))_{i=1}^n$ . We then compute

$$\begin{aligned} \mathbb{E}_N(\rho_n) &= \frac{1}{N} \sum_{j=1}^N \rho_{n,j}, & \mathbb{E}_N(\rho_n^{\text{rank}}) &= \frac{1}{N} \sum_{j=1}^N \rho_{n,j}^{\text{rank}}; & (2.22) \\ \sigma_N(\rho_n) &= \sqrt{\frac{1}{N-1} \sum_{j=1}^N (\rho_{n,j} - \mathbb{E}_N(\rho_n))^2}, \\ \sigma_N(\rho_n^{\text{rank}}) &= \sqrt{\frac{1}{N-1} \sum_{j=1}^N (\rho_{n,j}^{\text{rank}} - \mathbb{E}_N(\rho_n^{\text{rank}}))^2}. & (2.23) \end{aligned}$$

The results are presented in Table 1. We clearly see that  $\rho_n$  has a significant standard deviation, of which estimators are similar for different values of  $n$ . This means that in the limit as  $n \rightarrow \infty$ ,  $\rho_n$  is a random variable with a significant spread in its values, as stated in Theorem 2.1. Thus, by evaluating  $\rho_n$  for one

Model parameters	$N$	$10^3$	$10^2$		
	$n$	$10^2$	$10^3$	$10^4$	$10^5$
$\alpha = (1/2, 1/2, 0)$ $\beta = (0, 1/2, 1/2)$	$\mathbb{E}_N(\rho_n)$	0.4395	0.4365	0.4458	0.4067
	$\sigma_N(\rho_n)$	0.3399	0.3143	0.3175	0.3106
	$\mathbb{E}_N(\rho_n^{\text{rank}})$	0.4508	0.4485	0.4504	0.4519
	$\sigma_N(\rho_n^{\text{rank}})$	0.0922	0.0293	0.0091	0.0033
$\alpha = (1/2, 1/3, 1/6)$ $\beta = (1/6, 1/3, 1/2)$	$\mathbb{E}_N(\rho_n)$	0.8251	0.7986	0.8289	0.8070
	$\sigma_N(\rho_n)$	0.1151	0.1125	0.1108	0.1130
	$\mathbb{E}_N(\rho_n^{\text{rank}})$	0.8800	0.8850	0.8858	0.8856
	$\sigma_N(\rho_n^{\text{rank}})$	0.0248	0.0073	0.0023	0.0007
$\alpha = (1/2, -1/3, 1/6)$ $\beta = (1/6, 1/2, -1/3)$	$\mathbb{E}_N(\rho_n)$	-0.3052	-0.3386	-0.3670	-0.3203
	$\sigma_N(\rho_n)$	0.6087	0.5841	0.5592	0.5785
	$\mathbb{E}_N(\rho_n^{\text{rank}})$	-0.3448	-0.3513	-0.3503	-0.3517
	$\sigma_N(\rho_n^{\text{rank}})$	0.1202	0.0393	0.0120	0.0034

**Table 1.** Estimated mean and standard deviation of  $\rho_n$  and  $\rho_n^{\text{rank}}$  in  $N$  samples with linear dependence (2.10),  $\mathbb{P}(\xi > x) = x^{-1.1}$ ,  $x \geq 1$ .

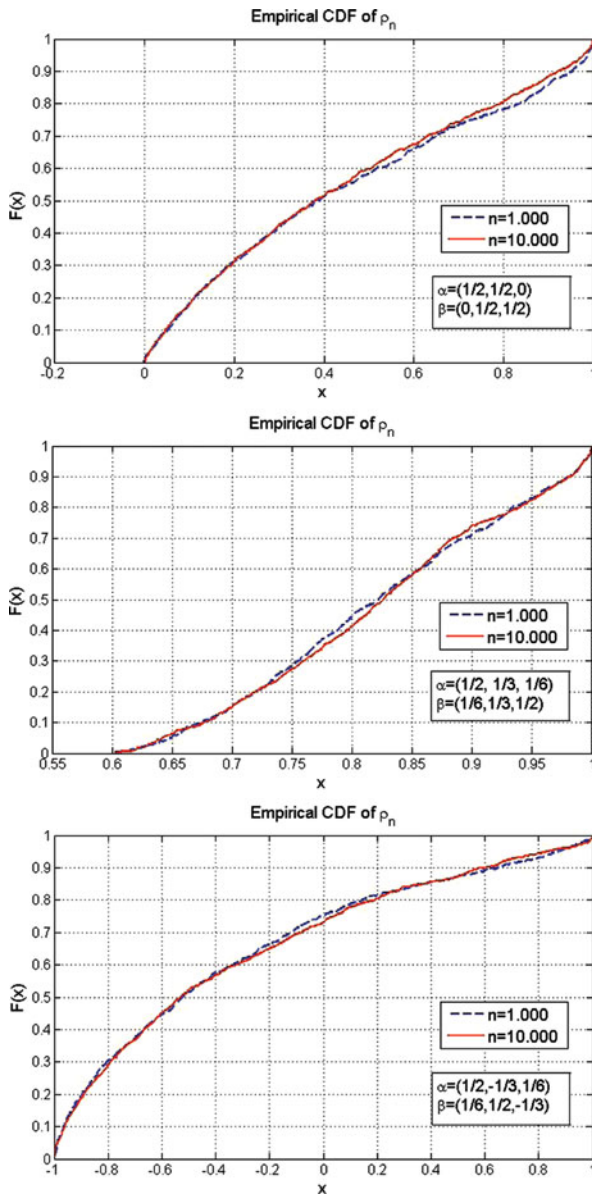
sample  $((X_i, Y_i))_{i=1}^n$  we will obtain a random number, even when  $n$  is huge. The convergence to a nontrivial distribution is directly seen in Figure 1 because the plots for the two values of  $n$  almost coincide. Note that in all cases, the density is fairly uniform, ensuring a comparable probability for all feasible values and rendering the value obtained in a specific realization even more uninformative.

However, from Table 1 we clearly see that the behavior of the rank correlation is exactly as we can expect from a good statistical estimator. The obtained average values are consistent whereas the standard deviation of  $\rho_n^{\text{rank}}$  decreases approximately  $1/\sqrt{n}$  as  $n$  grows large. Therefore,  $\rho_n^{\text{rank}}$  converges to a deterministic number.

#### 2.4. Sample Pearson's Correlation Coefficient for Nonnegative Variables

We proceed by investigating correlations between nonnegative heavy-tailed random variables. Our main result in this section shows that the correlation coefficient is asymptotically nonnegative:

**Theorem 2.3.** (Asymptotic nonnegativity of the sample Pearson's coefficient for positive r.v.'s). *Let  $((X_i, Y_i))_{i=1}^n$  be i.i.d. copies of nonnegative random variables  $(X, Y)$ , where  $X$*



**Figure 1.** The empirical distribution function  $F_N(x) = \mathbb{P}(\rho_n \leq x)$  for the  $N = 1,000$  observed values of  $\rho_n$  ( $n = 1,000, n = 10,000$ ), in the case of linear dependence (2.10).

$N$	$10^3$		$10^2$		
	10	$10^2$	$10^3$	$10^4$	$10^5$
$\mathbb{E}_N(\rho_n)$	-0.4833	-0.1363	-0.0342	-0.0077	-0.0015
$\sigma_N(\rho_n)$	0.1762	0.0821	0.0245	0.0064	0.0011
$\mathbb{E}_N(\rho_n^{\text{rank}})$	-0.6814	-0.4508	-0.4485	-0.4504	-0.4519
$\sigma_N(\rho_n^{\text{rank}})$	0.1580	0.0283	0.0082	0.0024	0.0007

**Table 2.** The mean and standard deviation of  $\rho_n$  and  $\rho_n^{\text{rank}}$  in  $N$  simulations of  $((X_i, Y_i))_{i=1}^n$ , where  $X = 2\xi I$ ,  $Y = 2\xi(1 - I)$ ,  $I$  is a Bernoulli(1/2) random variable,  $\mathbb{P}(\xi > x) = x^{-1.1}$ ,  $x \geq 1$ .

and  $Y$  satisfy

$$\mathbb{P}(X > x) = L_X(x)x^{-\gamma_X}, \quad \mathbb{P}(Y > y) = L_Y(y)y^{-\gamma_Y}, \quad x, y \geq 0, \quad (2.24)$$

with  $\gamma_X, \gamma_Y \in (0, 2)$ , so that  $\text{Var}(X) = \text{Var}(Y) = \infty$ . Then, any limit point of the sample Pearson correlation coefficient is nonnegative.

We illustrate Theorem 2.3 with a useful example. Let  $(\xi_i)_{i=1}^n$  be a sequence of i.i.d. random variables satisfying (2.11) for some  $\gamma \in (0, 2)$ , and where  $\xi \geq 0$  a.s. Let  $(X, Y) = (0, 2\xi)$  with probability 1/2 and  $(X, Y) = (2\xi, 0)$  with probability 1/2. Then,  $XY = 0$  a.s., whereas  $\mathbb{E}[X] = \mathbb{E}[Y] = \mathbb{E}[\xi]$  and  $\text{Var}(X) = \text{Var}(Y) = 2\mathbb{E}[\xi^2] - \mathbb{E}[\xi]^2 = 2\text{Var}(\xi) + \mathbb{E}[\xi]^2$ . By Theorem 2.3,  $\rho_n \xrightarrow{\mathbb{P}} 0$  when  $(\xi_i)_{i=1}^n$  is a sequence of i.i.d. nonnegative random variables satisfying (2.11) for some  $\gamma \in (0, 2)$ , which is not appropriate because  $(X, Y)$  are highly negatively dependent. When  $\gamma > 2$ , this anomaly does not arise, since, if  $\text{Var}(\xi) < \infty$ ,

$$\rho_n \xrightarrow{\mathbb{P}} \rho = -\frac{\mathbb{E}[\xi]^2}{2\text{Var}(\xi) + \mathbb{E}[\xi]^2} \in (-1, 0). \quad (2.25)$$

The asymptotics in (2.25) are quite reasonable, because the random variables  $(X, Y)$  are highly negatively dependent: When  $X > 0$ ,  $Y$  must be equal to 0, and vice versa.

Table 2 shows the empirical mean and standard deviation of the estimators  $\rho_n$  and  $\rho_n^{\text{rank}}$ . Here  $\mathbb{P}(\xi > x) = x^{-1.1}$ ,  $x \geq 1$ , as in Table 1. As predicted by Theorem 2.3, the sample correlation coefficient (assortativity) converges to zero as  $n$  grows large, whereas  $\rho_n^{\text{rank}}$  consistently shows a clear negative dependence, and the precision of the estimator improves as  $n \rightarrow \infty$ . This explains why strong disassortativity is not observed in large samples of nonnegative power-law data.

We next prove Theorem 2.3:

**Proof of Theorem 2.3.** Clearly  $\sum_{i=1}^n X_i Y_i \geq 0$  when  $X_i \geq 0, Y_i \geq 0$ , so that

$$\rho_n \geq -\frac{\frac{1}{n-1} \sum_{i=1}^n \bar{X}_n \bar{Y}_n}{S_n(X)S_n(Y)} = -\frac{n}{n-1} \frac{\bar{X}_n}{S_n(X)} \frac{\bar{Y}_n}{S_n(Y)}.$$

It remains to be shown that if  $\text{Var}(X) = \infty$ , then  $\bar{X}_n/S_n(X) \xrightarrow{\mathbb{P}} 0$ . Indeed, if  $\gamma \in (1, 2)$  then  $\bar{X}_n \xrightarrow{\mathbb{P}} \mathbb{E}[X] < \infty$  by the strong law of large numbers. When  $\gamma \in (0, 1]$ , instead, then  $X$  is in the domain of attraction of a  $\gamma$  stable random variable, hence  $\bar{X}_n$ , loosely speaking, scales as  $n^{1/\gamma_X - 1}$ . Further, from (2.24) and Lemma 2.2 it follows that  $S_n(X)$  scales as  $n^{2/\gamma_X - 1}$ , in particular,  $\bar{X}_n/S_n(X) \xrightarrow{\mathbb{P}} 0$  for all  $\gamma \in (0, 2)$ .  $\square$

### 3. Applications to Networks

In real-world networks it is particularly important to measure degree-degree dependencies for neighboring vertices. We refer to [Newman 10] for an extensive introduction to networks, their empirical properties and models for them. In Section 3.1, we start with the formal definition of Pearson’s correlation coefficient (which was termed the *assortativity coefficient* in [Newman 02]), and Spearman’s rho in the network context. Next, in Section 3.2 we show that all limit points of Pearson’s coefficients for sequences of growing scale-free random graphs with power-law exponent  $\gamma < 3$  are nonnegative, a result that is similar in spirit to Theorem 2.3. In Section 3.3, we state general convergence conditions for both Pearson’s correlation coefficient as well as Spearman’s rho.

#### 3.1. Definitions and Notations

We start by introducing some notation. Let  $G = (V, E)$  be an undirected random graph. For a *directed* edge  $e = (u, v)$ , we write  $\underline{e} = u, \bar{e} = v$  and we denote the set of directed edges in  $E$  by  $E'$  (so that  $|E'| = 2|E|$ ), and  $D_v$  is the degree of vertex  $v \in V$ . In general,  $D_v$  is a random variable.

The assortativity coefficient of  $G$  is equal to (see, e.g., [Newman 02, (4)])

$$\rho(G) = \frac{\frac{1}{|E'|} \sum_{(u,v) \in E'} D_u D_v - \left( \frac{1}{|E'|} \sum_{(u,v) \in E'} \frac{1}{2} (D_u + D_v) \right)^2}{\frac{1}{|E'|} \sum_{(u,v) \in E'} \frac{1}{2} (D_u^2 + D_v^2) - \left( \frac{1}{|E'|} \sum_{(u,v) \in E'} \frac{1}{2} (D_u + D_v) \right)^2}. \quad (3.1)$$

Note that the assortativity coefficient in (3.1) is equal to the sample correlation coefficient, where  $((D_u, D_v))_{(u,v) \in E'}$  represents a sequence of nonnegative random variables, as studied in Theorem 2.3. However,  $((D_u, D_v))_{(u,v) \in E'}$  is not independent, so that we may not immediately apply the previous theory. Theorem 3.1 is the analogue of Theorem 2.3 in the network context, and we give a formal proof of it below.

Let us now introduce Spearman’s rho in  $G$  that we denote by  $\rho^{\text{rank}}(G)$ . In accordance to the original definition of Spearman’s rho,  $\rho^{\text{rank}}(G)$  is the correlation coefficient of the sequence of random variables  $(R_e, R_{\bar{e}})$ , where  $e$  is a uniformly chosen directed edge  $(u, v)$  from  $E'_n$ . We let  $R_{\underline{e}}$  and  $R_{\bar{e}}$  be the rank of, respectively,  $D_{\underline{e}} + U_e$  and  $D_{\bar{e}} + U'_e$  in the sequences  $(D_{\underline{e}} + U_e)_{e \in E'_n}$  and  $(D_{\bar{e}} + U'_e)_{e \in E'_n}$ . Here, as discussed earlier,  $(U_e)_{e \in E'_n}$  and  $(U'_e)_{e \in E'_n}$  are i.i.d. sequences of uniform  $(0, 1)$  random variables. Then, Spearman’s rank correlation coefficient is defined as follows:

$$\rho^{\text{rank}}(G) = \frac{\frac{1}{|E'|} \sum_{e \in E'} R_e R_{\bar{e}} - (|E'| + 1)^2 / 4}{(|E'|^2 - 1) / 12}. \tag{3.2}$$

### 3.2. No Disassortative Scale-Free Random Graph Sequences

We compute that

$$\begin{aligned} \frac{1}{|E'|} \sum_{(u,v) \in E'} \frac{1}{2} (D_u + D_v) &= \frac{1}{|E'|} \sum_{v \in V} D_v^2, \\ \frac{1}{|E'|} \sum_{(u,v) \in E'} \frac{1}{2} (D_u^2 + D_v^2) &= \frac{1}{|E'|} \sum_{v \in V} D_v^3. \end{aligned} \tag{3.3}$$

Thus,  $\rho(G)$  can be written as

$$\rho(G) = \frac{\sum_{(u,v) \in E'} D_u D_v - \frac{1}{|E'|} \left( \sum_{v \in V} D_v^2 \right)^2}{\sum_{v \in V} D_v^3 - \frac{1}{|E'|} \left( \sum_{v \in V} D_v^2 \right)^2}. \tag{3.4}$$

Consider a sequence of graphs  $(G_n)_{n \geq 1}$ , where  $G_n = (V_n, E_n)$  and  $n$  denotes the number of vertices  $n = |V_n|$  in the graph. Since many real-world networks are quite large, we are interested in the behavior of  $\rho(G_n)$  as  $n \rightarrow \infty$ . Note that this discussion applies both to sequences of real-world networks of increasing size, as well as to graph sequences of random graphs. We start by generalizing Theorem 2.3 to this setting:



**Theorem 3.1. (Asymptotic nonnegativity of Pearson’s coefficient in scale-free graphs).** *Let  $(G_n)_{n \geq 1}$  be a sequence of graphs of size  $n$  satisfying that there exist  $\gamma \in (1, 3)$  and  $0 < c < C < \infty$  such that  $cn \leq |E| \leq Cn$ ,  $cn^{1/\gamma} \leq \max_{v \in V_n} D_v \leq Cn^{1/\gamma}$  and  $cn^{(2/\gamma)\vee 1} \leq \sum_{v \in V_n} D_v^2 \leq Cn^{(2/\gamma)\vee 1}$ . Then, any limit point of Pearson’s correlation coefficient  $\rho(G_n)$  is non-negative.*

In the next section, we give several examples where Theorem 3.1 applies and yields results that are not sensible. The powerful feature of Theorem 3.1 is that it applies to *all* graphs, not just realizations of certain random graphs.

**Proof.** We note that  $D_v \geq 0$  for every  $v \in V$ , so that, from (3.4)

$$\rho(G_n) \geq \rho^-(G_n) \equiv -\frac{\frac{1}{|E'|} \left( \sum_{v \in V} D_v^2 \right)^2}{\sum_{v \in V} D_v^3 - \frac{1}{|E'|} \left( \sum_{v \in V} D_v^2 \right)^2}. \tag{3.5}$$

By assumption,  $\sum_{v \in V} D_v^3 \geq (\max_{v \in [n]} D_v)^3 \geq c^3 n^{3/\gamma}$ , whereas  $\frac{1}{|E'|} \left( \sum_{v \in V} D_v^2 \right)^2 \leq (C^2/c)n^{2(2/\gamma \vee 1)-1} = (C^2/c)n^{[(4/\gamma-1)\vee 1]}$ . Since  $\gamma \in (1, 3)$  we have  $(4/\gamma - 1) \vee 1 < 3/\gamma$ , so that

$$\frac{\sum_{v \in V} D_v^3}{\frac{1}{|E'|} \left( \sum_{v \in V} D_v^2 \right)^2} \rightarrow \infty.$$

Hence,  $\rho^-(G_n) \rightarrow 0$  as  $n \rightarrow \infty$ . This proves the claim. □

In the literature, many examples are reported of real-world networks where the degree distribution closely follows a power law with  $\gamma$  in  $(1, 3)$ , see e.g., [Albert and Barabási 02, Table I] or [Newman 03b, Table I]. Let  $D$  be such a power-law random variable, and denote  $\mu_p = \mathbb{E}[D^p]$  for  $p \in (0, \gamma)$ . In that case, one can expect that

$$|E'| = \sum_{v \in V} D_v \sim \mu_1 n,$$

while  $\max_{v \in V} D_v \sim n^{1/\gamma}$ , and

$$\frac{1}{n} \sum_{v \in V} D_v^p \sim \begin{cases} \mu_p & \text{when } \gamma > p, \\ C_p n^{p/\gamma-1} & \text{when } \gamma < p. \end{cases} \tag{3.6}$$

Of course, the convergence in (3.6) depends sensitively on the occurrence of large degrees. However, intuitively it can be explained as follows. When

$$\frac{1}{n} \sum_{v \in V} \mathbb{1}_{\{D_v \geq k\}} = C' k^{-\gamma} (1 + o(1))$$

for all  $k$  for which  $k^{-\gamma} \gg 1/n$  so that  $k \ll n^{1/\gamma}$ , then

$$\frac{1}{n} \sum_{v \in V} D_v^p = \sum_{k \geq 1} (k^p - (k-1)^p) \frac{1}{n} \sum_{v \in V} \mathbb{1}_{\{D_v \geq k\}} \approx C'' \sum_{k=1}^{n^{1/\gamma}} k^{p-1-\gamma} = C_p n^{p/\gamma-1},$$

where  $C''$  and  $C_p$  are appropriately chosen constants. In particular, the conditions of Theorem 3.1 hold and  $\rho^-(G_n) \rightarrow 0$  when  $\gamma < 3$ . Thus, the asymptotic degree-degree correlation of the graph sequence  $(G_n)_{n \geq 1}$  is nonnegative. As a result, when the power-law exponent satisfies  $\gamma < 3$  there exist no scale-free graph sequences that will be identified as disassortative by Pearson's coefficient. We next state a general theorem that allows us to identify the limit of Spearman's rho and Pearson's coefficient for many random graph models.

### 3.3. Convergence Conditions for Degree-Degree Dependency Measures

Let  $(G_n)_{n \geq 1}$  be again a sequence of graphs of size  $n$ , where  $G_n = (V_n, E_n)$ ,  $|V_n| = n$ . We write  $\mathbb{E}_n$  for the conditional expectation given the graph  $G_n$  (which in itself is random, so that we are *not* taking the expectation w.r.t.  $G_n$ ). Consider a random vector  $(X, Y) = (D_e, D_{\bar{e}})$ , where  $e$  is chosen uniformly at random from  $E'$ . Recall that for a discrete random variable  $X$ ,  $F_X$  denotes its cumulative distribution function, and  $F_X^*$  denotes the cumulative distribution function of  $X^* = X + U$ , where  $U$  is an independent uniform random variable on  $(0, 1)$ . Then  $F_X^*(X^*)$  has a uniform distribution on  $(0, 1)$ , see (2.8). Our main result to identify the limits of Spearman's rho as given by (3.2) and Pearson's coefficient is the following theorem:

**Theorem 3.2. (Convergence criteria for degree-degree dependency measures).** *Let  $(G_n)_{n \geq 1}$  be a sequence of random graphs of size  $n$ , where  $G_n = (V_n, E_n)$ ,  $|V_n| = n$ . Let  $(X_n, Y_n)$  be the degrees on both sides of a uniform directed edge  $e \in E'_n$ . Suppose that for every bounded continuous  $h: \mathbb{R}^2 \rightarrow \mathbb{R}$ ,*

$$\mathbb{E}_n[h(X_n, Y_n)] \xrightarrow{\mathbb{P}} \mathbb{E}[h(X, Y)], \quad (3.7)$$

where the r.h.s. is nonrandom. Then

(a)

$$\rho^{\text{rank}}(G_n) \xrightarrow{\mathbb{P}} 12\mathbb{E}(F_X^*(X^*)F_Y^*(Y^*)) - 3 = \rho^{\text{rank}}, \tag{3.8}$$

where  $X^* = X + U, Y^* = Y + U'U$  and  $U'$  are independent random variables on  $(0,1)$ , also independent of  $X$  and  $Y$ , and  $F_{X^*}(\cdot)$  is the cumulative distribution function of  $X^*$ ;

(b) when we further suppose that  $\mathbb{E}_n[X_n^2] \xrightarrow{\mathbb{P}} \mathbb{E}[X^2] < \infty$ , and  $\text{Var}(X) > 0$ , then also

$$\rho(G_n) \xrightarrow{\mathbb{P}} \rho = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}. \tag{3.9}$$

We remark that when  $G_n$  is a random graph, then  $\rho^{\text{rank}}(G_n)$  and  $\rho(G_n)$  are random variables. Equation (3.7) implies that the *distribution* of the degrees on either side of an edge converges in probability to a deterministic limit, which can be interpreted as the statement that the degree distribution converges to a deterministic limit. The limits of  $\rho^{\text{rank}}(G_n)$  and  $\rho(G_n)$  depend only on the limiting degree distribution, where  $\rho^{\text{rank}}(G_n)$  *always* converges, while  $\rho(G_n)$  can be proved to converge only when its limit is well defined. We further note that (3.7) is equivalent to showing that

$$\#\{e = (u, v) \in E'_n : (D_u, D_v) = (k, l)\} / |E'_n| \xrightarrow{\mathbb{P}} \mathbb{P}(X = k, Y = l). \tag{3.10}$$

Condition (3.10) will be simpler to verify in practice. We emphasize that we study *undirected* graphs but we work with *directed* edges  $e = (u, v)$ , which we vary over the whole set of edges, in such a way that  $(u, v)$  and  $(v, u)$  contribute as different edges. In particular, the marginal distributions of  $X_n$  and  $Y_n$ , and consequently of  $X$  and  $Y$ , are the same. We next prove Theorem 3.2:

**Proof.** We start with part (a). The sequence  $(R_{\underline{e}}/|E'_n|, R_{\bar{e}}/|E'_n|)$  is a bounded sequence of two-dimensional random variables. Let  $F_{n,X}$  denote the empirical cumulative distribution function of  $(D_{\underline{e}})_{e \in E'_n}$  (which equals that of  $(D_{\bar{e}})_{e \in E'_n}$ ), and let  $F_{n,X}^*$  denote the empirical cumulative distribution functions of  $(D_{\underline{e}} + U_e)_{e \in E'_n}$  (which equals that of  $(D_{\bar{e}} + U'_e)_{e \in E'_n}$ ), where  $(U_e)_{e \in E'_n}, (U'_e)_{e \in E'_n}$  are independent sequences of i.i.d uniform  $(0, 1)$  random variables. Then, we can

rewrite, with  $\ell_n = |E'_n|$ ,

$$(R_{\underline{e}}, R_{\bar{e}}) = ((\lceil \ell_n F_{n,x}^* (D_{\underline{e}} + U_e) \rceil, \lceil \ell_n F_{n,x}^* (D_{\bar{e}} + U'_e) \rceil). \tag{3.11}$$

In particular,

$$(R_{\underline{e}}/\ell_n, R_{\bar{e}}/\ell_n) = (\lceil \ell_n F_{n,x}^* (D_{\underline{e}} + U_e) \rceil/\ell_n, \lceil \ell_n F_{n,x}^* (D_{\bar{e}} + U'_e) \rceil/\ell_n). \tag{3.12}$$

Thus,

$$(R_{\underline{e}}/\ell_n, R_{\bar{e}}/\ell_n) = (F_{n,x}^* (D_{\underline{e}} + U_e), F_{n,x}^* (D_{\bar{e}} + U'_e)) + O(1/\ell_n). \tag{3.13}$$

By (3.7), the fact that  $X_n \xrightarrow{d} X$  and the fact that  $F_x^*$  is continuous,  $F_{n,x}^*(x) \xrightarrow{\mathbb{P}} F_x^*(x)$  for every  $x \geq 0$ . Moreover, we claim that this convergence holds uniformly in  $x$ , i.e.,  $\sup_{x \in \mathbb{R}} |F_{n,x}^*(x) - F_x^*(x)| \xrightarrow{\mathbb{P}} 0$ . To see this, note that (3.7) implies that the distribution functions of  $X_n$  and  $Y_n$  converge to those of  $X$  and  $Y$ . Since all these random variables take on only integer values, this convergence is *uniform*, i.e.,  $\sup_{k \geq 0} |F_{n,x}(k) - F_x(k)| \xrightarrow{\mathbb{P}} 0$ . We obtain  $F_{n,x}^*$  by linearly interpolating between  $F_{n,x}(k-1)$  and  $F_{n,x}(k)$  for every  $k$ , so also  $F_{n,x}^*$  converges uniformly, as we claimed.

By this uniform convergence, for every bounded continuous function  $g: [0, 1]^2 \rightarrow \mathbb{R}$ ,

$$\begin{aligned} \mathbb{E}_n[g(R_{\underline{e}}/\ell_n, R_{\bar{e}}/\ell_n)] &= \mathbb{E}_n[g(F_{n,x}^*(D_{\underline{e}} + U_e), F_{n,x}^*(D_{\bar{e}} + U'_e))] \\ &= \mathbb{E}_n[g(F_x^*(D_{\underline{e}} + U_e), F_x^*(D_{\bar{e}} + U'_e))] + o_{\mathbb{P}}(1) \\ &= \mathbb{E}_n[g(F_x^*(X_n + U), F_x^*(Y_n + U'))] + o_{\mathbb{P}}(1) \\ &\xrightarrow{\mathbb{P}} \mathbb{E}[g(F_x^*(X + U), F_x^*(Y + U'))] = \mathbb{E}[g(F_x^*(X^*), F_x^*(Y^*))], \end{aligned} \tag{3.14}$$

again by (3.7) and the fact that  $(x, y) \mapsto \mathbb{E}[g(F_x^*(x + U), F_x^*(y + U'))]$  is continuous and bounded. Applying this to  $g(x, y) = xy$ ,  $g(x, y) = x^2$  and  $g(x, y) = y^2$  yields the required convergence. Moreover, since  $F_x^*(X^*)$  and  $F_x^*(Y^*)$  are uniform random variables,  $\text{Var}(F_x^*(X^*)) = \text{Var}(F_x^*(Y^*)) = 1/12$ . This completes the proof of convergence in (a). The equality in (a) is just [Mesfioui and Tajar 05, Proposition 3.1], see (2.9).

For part (b), we note that

$$\rho(G_n) = \frac{\text{Cov}_n(X_n, Y_n)}{\text{Var}_n(X_n)}. \tag{3.15}$$

Since  $\mathbb{E}_n[X_n^2] \xrightarrow{\mathbb{P}} \mathbb{E}[X^2] < \infty$ , also  $\mathbb{E}_n[X_n] \xrightarrow{\mathbb{P}} \mathbb{E}[X] < \infty$ , so that  $\text{Var}_n(X_n) \xrightarrow{\mathbb{P}} \text{Var}(X)$ . Since these limits are positive, by Slutsky's theorem,

$$\rho(G_n) = \frac{\text{Cov}_n(X_n, Y_n)}{\text{Var}(X)}(1 + o_{\mathbb{P}}(1)). \tag{3.16}$$

Furthermore, the random variables  $(X_n Y_n)_{n \geq 1}$  converge in distribution, and are uniformly integrable (since both  $(X_n^2)_{n \geq 1}$  and  $(Y_n^2)_{n \geq 1}$  are, which again follows from the fact that  $\mathbb{E}_n[X_n^2] \xrightarrow{\mathbb{F}} \mathbb{E}[X^2] < \infty$  and the fact that  $X_n$  and  $Y_n$  have the same marginals). Therefore, also  $\mathbb{E}_n[X_n Y_n] \xrightarrow{\mathbb{F}} \mathbb{E}[XY]$ , so that the convergence follows.  $\square$

## 4. Random Graph Examples

In this section we consider four random graph models to highlight our result: the configuration model, the configuration model with intermediate vertices, the preferential attachment model, and a model of complete bipartite random graphs. In Section 5, we present the numerical results for these models.

### 4.1. The Configuration Model

The *configuration model* (CM) was invented by Bollobás in [Bollobás 80] and inspired by [Bender and Canfield 78]. Its connectivity structure was first studied by Molloy and Reed [Molloy and Reed 95, 98]. It was popularized by Newman, Srogoatz and Watts [Newman et al. 01], who realized that it is a useful and simple model for real-world networks.

Given a *degree sequence*, namely a sequence of  $n$  positive integers  $\mathbf{d} = (d_1, d_2, \dots, d_n)$  with  $\ell_n = \sum_{i \in [n]} d_i$  assumed to be even, the configuration model (CM) on  $n$  vertices and degree sequence  $\mathbf{d}$  is constructed as follows. Start with  $n$  vertices labeled  $1, 2, \dots, n$ , and  $d_v$  halfedges adjacent to vertex  $v$ . The graph is constructed by randomly pairing each halfedge to some other halfedge to form an edge. Number the halfedges from 1 to  $\ell_n$  in some arbitrary order. Then, at each step, two halfedges that are not already paired are chosen uniformly at random among all the unpaired halfedges and are paired to form a single edge in the graph. These halfedges are removed from the list of unpaired halfedges. We continue with this procedure of choosing and pairing two unpaired halfedges until all the halfedges are paired. In the resulting graph  $G_n = (V_n, E_n)$  we have  $|V_n| = n$ ,  $\ell_n = 2|E_n|$ . Although self-loops and double edges may occur, these become rare as  $n \rightarrow \infty$  (see e.g. [Bollobás 01] or [Janson 09] for more precise results in this direction). In the analysis, we keep the self-loops and multiple edges, so that  $\ell_n = |E'_n|$ . In the numerical simulation we also consider the case where the self-loops are removed, and we collapse multiple edges to a single edge. As we will see in the simulations, these two cases are qualitatively similar.

We investigate the CM where the degrees are i.i.d. random variables, and note that the probability that two vertices  $u$  and  $v$  are directly connected is close to  $d_u d_v / \ell_n$ . Because this is of product form in  $u$  and  $v$ , the degrees at either end of an

edge are close to being independent, and in fact are asymptotically independent. Therefore, one expects the assortativity coefficient of the configuration model to converge to 0 in probability, irrespective of the degree distribution.

We now make this argument precise. We make the following assumptions on our degree sequence  $(d_v)_{v \in V_n}$ :

**Condition 4.1. (Degree regularity).**

(a) There exists a probability distribution  $(p_k)_{k \geq 0}$  such that  $n_k/n \rightarrow p_k$  for every  $k \geq 1$ , where  $n_k = \#\{v: d_v = k\}$  denotes the number of vertices of degree  $k$ .

(b)  $\mathbb{E}[D_{(n)}] \rightarrow \mathbb{E}[D]$ , where  $\mathbb{P}(D_{(n)} = k) = n_k/n$  and  $\mathbb{P}(D = k) = p_k$ .

See [van der Hofstad 13, Chapter 7] for an extensive discussion of the CM under Condition 4.1.

**Theorem 4.2. (Convergence of the degree-degree dependency measures for CM).** *Let  $(G_n)_{n \geq 1}$  be a sequence of configuration models of size  $n$ , for which the degree sequence  $(d_v)_{v \in V_n}$  satisfies Condition 4.1. Then*

$$\rho^{\text{rank}}(G_n) \xrightarrow{\mathbb{P}} 0,$$

and

$$\rho(G_n) \xrightarrow{\mathbb{P}} 0.$$

**Proof.** We apply Theorem 3.2, and we start by investigating (3.10). We note that a uniform edge can be constructed by taking two halfedges uniformly at random. Indeed, we can first draw the first half-edge uniformly at random, and this will be paired to another half-edge uniformly at random by construction of the CM. We perform a second moment argument on  $N_{k,l} = \#\{e = (u, v) \in E'_n : (d_u, d_v) = (k, l)\}$ , and will prove that

$$N_{k,l}/\ell_n \xrightarrow{\mathbb{P}} \frac{kp_k}{\mathbb{E}[D]} \frac{lp_l}{\mathbb{E}[D]}.$$

For this, it suffices to prove that

$$\mathbb{E}[N_{k,l}]/\ell_n \rightarrow \frac{kp_k}{\mathbb{E}[D]} \frac{lp_l}{\mathbb{E}[D]}, \quad \mathbb{E}[N_{k,l}^2]/\ell_n^2 \rightarrow \left( \frac{kp_k}{\mathbb{E}[D]} \frac{lp_l}{\mathbb{E}[D]} \right)^2,$$

since, then,  $\text{Var}(N_{k,l}/\ell_n) = o(1)$ .

We note that

$$\mathbb{E}[N_{k,l}] = \frac{kl n_k n_l}{\ell_n - 1},$$

where  $\ell_n = \sum_{v \in V_n} d_v = 2|E_n|$  and  $n_k = \#\{v : d_v = k\}$  is the number of vertices with degree  $k$ . Therefore, also using that  $\ell_n = n\mathbb{E}[D_{(n)}]$ , Condition 4.1 implies that

$$\mathbb{E}[N_{k,l}]/\ell_n \rightarrow \frac{kp_k}{\mathbb{E}[D]} \frac{lp_l}{\mathbb{E}[D]}.$$

Further,

$$\mathbb{E}[N_{k,l}^2]/\ell_n^2 = \frac{1}{\ell_n^2} \sum_{(u_1, v_1), (u_2, v_2)} \mathbb{P}(d_{u_1} = k, d_{v_1} = l, d_{u_2} = k, d_{v_2} = l).$$

There are four different cases, depending on  $a = \#\{u_1, u_2, v_1, v_2\}$ . When  $a = 4$ , the contribution is

$$\frac{k^2 n_k (n_k - 1) l^2 n_l (n_l - 1)}{\ell_n^2 (\ell_n - 1) (\ell_n - 3)} = \frac{(kn_k ln_l)^2}{\ell_n^4} (1 + O(1/n)) \rightarrow \left( \frac{kp_k}{\mathbb{E}[D]} \frac{lp_l}{\mathbb{E}[D]} \right)^2.$$

Therefore, we are shown that the contributions due to  $a \leq 3$  vanish.

When  $a = 3$ , either one of the edges  $(u_1, v_1)$  and  $(u_2, v_2)$  is a self-loop, while the other joins two other vertices (which only contributes when  $k = l$ ), or both edges start in the same vertex  $v$ , so that this contribution is, at most,

$$\frac{k^2 n_k (n_k - 1) l^2 n_l}{\ell_n^2 (\ell_n - 1) (\ell_n - 3)} = O(1/n) = o(1).$$

When  $a = 2$ , similar computations show that the contribution is, at most,  $O(1/n^2)$ . When  $a = 1$ , the edges  $(u_1, v_1)$  and  $(u_2, v_2)$  are self-loops from the same vertex  $v$ , so this contributes only when  $k = l$ , and then, at most,

$$\frac{k(k-1)(k-2)(k-3)n_k}{\ell_n^2 (\ell_n - 1) (\ell_n - 3)} = O(1/n^3) = o(1).$$

We conclude that (3.10) holds with

$$\mathbb{P}(X = k, Y = l) = \frac{kp_k}{\mathbb{E}[D]} \frac{lp_l}{\mathbb{E}[D]}.$$

In particular,  $X$  and  $Y$  are independent, so that  $\rho^{\text{rank}} = 0$ . This proves the first part of Theorem 4.2.

For the second part, we note that when the degrees  $(d_v)_{v \in V_n}$  are fixed, the only random part in  $\rho(G_n)$  is

$$M_n = \frac{1}{\ell_n} \sum_{e \in E'_n} d_{\underline{e}} d_{\bar{e}}.$$

We perform a second moment method on this quantity. We use an edge  $e$  that is a pair of two specified halfedges incident to two specific vertices. Thus, we can denote  $e$  by  $\underline{e} = (u, s), \bar{e} = (v, t)$ , where  $u, v$  are the vertices to which the specific

halfedges are incident, while  $s \in \{1, \dots, d_u\}$  is the label of the halfedge incident to vertex  $u$ , and  $t \in \{1, \dots, d_v\}$  is the label of the halfedge incident to vertex  $v$ , which are paired together. The probability of pairing them together equals  $1/(\ell_n - 1)$ . Therefore,

$$\mathbb{E}[M_n] = \frac{1}{\ell_n} \sum_{u,v,s,t} \frac{d_u d_v}{\ell_n - 1} = \sum_{u,v \in V_n} d_u^2 d_v^2 / \ell_n (\ell_n - 1) = \sum_{u,v \in V_n} d_u^2 d_v^2 / \ell_n^2 (1 + O(1/n)),$$

where we note that we count multiple edges as frequently as they occur. Further, and in a similar way,

$$\mathbb{E}[M_n^2] = (1 + o(1)) \sum_{u,v,u',v' \in V_n} d_u^2 d_u'^2 d_v^2 d_v'^2 / \ell_n^4,$$

so that

$$\frac{M_n}{\left(\sum_{v \in V_n} d_v^2 / \ell_n\right)^2} \xrightarrow{\mathbb{P}} 1.$$

In particular,

$$\rho(G_n) = \frac{M_n - \left(\sum_{u,v \in V_n} d_u^2 / \ell_n\right)^2}{\sum_{u \in V_n} d_u^3 / \ell_n - \left(\sum_{u \in V_n} d_u^2 / \ell_n\right)^2} \xrightarrow{\mathbb{P}} 0,$$

both when  $\sum_{u \in V_n} d_u^3 / \ell_n \gg (\sum_{u \in V_n} d_u^2 / \ell_n)^2$ , as well as when  $\sum_{u \in V_n} d_u^3 / \ell_n = \Theta(\sum_{u \in V_n} d_u^2 / \ell_n)^2$ .  $\square$

## 4.2. Configuration Model with Intermediate Vertices

We now give an example of a strongly disassortative graph to demonstrate that  $\rho(G_n)$  fails to capture obvious negative degree-degree dependencies when the degree distribution is heavy tailed. In order to do that we adapt the configuration model slightly by replacing every edge by two edges that meet at a middle vertex. Denote this graph by  $\tilde{G}_n = (\tilde{V}_n, \tilde{E}_n)$ , while the configuration model is  $G_n = (V_n, E_n)$ . In this model, there are  $n + \ell_n/2$  vertices and  $|\tilde{E}'_n| = 2\ell_n$  directed edges. For  $(u, v) \in \tilde{E}'_n$ , the degree of either vertex  $u$  or vertex  $v$  equals 2, and the degree of the other vertex in the edge is equal to  $d_s$ , where  $s$  is the unique vertex in the original configuration model that corresponds to  $u$  or  $v$ .

**Theorem 4.3. (Convergence of degree-degree dependency measures for CM with intermediate vertices).** *Let  $(\tilde{G}_n)_{n \geq 1}$  be a sequence of configuration models with intermediate vertices, where*



the degree sequence  $(d_v)_{v \in V_n}$  satisfies Condition 4.1. Then

$$\begin{aligned} \rho^{\text{rank}}(\bar{G}_n) &\xrightarrow{\mathbb{P}} 12\mathbb{E}(F_X(X)F_X(Y)) - 3 \\ &= -\frac{3}{4} + 3\left(\tilde{p}_1 + \frac{1}{2}\tilde{p}_2\right)\left(1 - \tilde{p}_1 - \frac{1}{2}\tilde{p}_2\right), \end{aligned} \tag{4.1}$$

where  $(X, Y) = (2I + (1 - I)\tilde{D}_1, 2(1 - I) + I\tilde{D}_2)$  with  $\tilde{D}_1, \tilde{D}_2$  i.i.d. random variables with  $\mathbb{P}(\tilde{D} = k) = kp_k/\mathbb{E}[D] := \tilde{p}_k$  and  $I$  an independent Bernoulli(1/2) random variable. Further,

$$\rho(G_n) \xrightarrow{\mathbb{P}} \begin{cases} \frac{\text{Cov}(X, Y)}{\text{Var}(X)} & \text{if } \mathbb{E}[D_{(n)}^3] \rightarrow \mathbb{E}[D^3] < \infty; \\ 0 & \text{if } \mathbb{E}[D_{(n)}^3] \rightarrow \infty, \end{cases}$$

and, for  $\mathbb{E}[D_{(n)}^3] \rightarrow \mathbb{E}[D^3] < \infty$ , and writing  $\mu_p = \mathbb{E}[D^p]$ ,

$$\frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{2\mu_2/\mu_1 - (1 + \mu_2/(2\mu_1))^2}{(2 + \mu_3/(2\mu_1)) - (1 + \mu_2/(2\mu_1))^2} < 0.$$

The fact that the degree-degree correlation is negative is quite reasonable, because in this model, vertices of high degree are connected to only vertices of degree 2, so that there is a negative dependence between the degrees at either end of an edge. When  $\mathbb{E}[D_{(n)}^3] \rightarrow \infty$ , however,  $\rho(\bar{G}_n) \xrightarrow{\mathbb{P}} 0$ , which is inappropriate, because the negative dependence of the degrees persists.

**Proof.** The first part follows directly from Theorem 3.2, since the collection of values  $(\bar{d}_e, \bar{d}_{\bar{e}})_{e \in \bar{E}'_n}$  depends only on the degrees  $(d_v)_{v \in V_n}$  and

$$\#\{e: \bar{d}_e = l, \bar{d}_{\bar{e}} = k\} / |\bar{E}'_n| = (kn_k\delta_{2,l} + ln_l\delta_{2,k} - 2n_2\mathbb{1}_{\{k=l=2\}}) / (2\ell_n),$$

which converges to  $\mathbb{P}(X = k, Y = 2)$ . Now, consider the possible values of  $X$ , and notice that

$$\mathbb{P}(X = 1) = \tilde{p}_1/2, \tag{4.2}$$

$$\mathbb{P}(X = 2) = 1/2 + \tilde{p}_2/2, \tag{4.3}$$

$$\mathbb{P}(X \geq 3) = 1/2 - \tilde{p}_1/2 - \tilde{p}_2/2. \tag{4.4}$$

Then we obtain

$$F_X^*(x + U) = \begin{cases} \frac{1}{2}\tilde{p}_1U, & \text{if } x = 1, \\ \frac{\tilde{p}_1}{2} + \left(\frac{\tilde{p}_2}{2} + \frac{1}{2}\right)U, & \text{if } x = 2, \\ \frac{1}{2} + \sum_{k=1}^{x-1} \frac{\tilde{p}_k}{2} + \frac{\tilde{p}_x}{2}U, & \text{if } x \geq 3. \end{cases} \tag{4.5}$$

Since either  $X$  or  $Y$  equals 2 and corresponds to the intermediate node, we further condition on  $\tilde{D}$ :

$$\begin{aligned} \mathbb{E}(F_x^*(X^*)F_x^*(Y^*)) &= \mathbb{E}(F_x^*(\tilde{D} + U)F_x^*(2 + U')) \\ &= \mathbb{E}(F_x^*(2 + U')) \\ &\times [(\mathbb{E}(F_x^*(1 + U))\mathbb{P}(\tilde{D} = 1) + \mathbb{E}(F_x^*(2 + U))\mathbb{P}(\tilde{D} = 2) \\ &+ \mathbb{E}(\tilde{D} + U|\tilde{D} \geq 3)\mathbb{P}(\tilde{D} \geq 3)]. \end{aligned} \quad (4.6)$$

Now, using (4.5) and substituting (4.2–4.4) from the last expression, we readily obtain

$$\begin{aligned} \mathbb{E}(F_x^*(X^*)F_x^*(Y^*)) &= \left( \frac{\tilde{p}_1}{2} + \frac{\tilde{p}_2}{4} + \frac{1}{4} \right) \\ &\times \left[ \frac{1}{4}(\tilde{p}_1)^2 + \left( \frac{\tilde{p}_1}{2} + \frac{\tilde{p}_2}{4} + \frac{1}{4} \right)\tilde{p}_2 \right. \\ &\left. + \left( \frac{\tilde{p}_1}{4} + \frac{\tilde{p}_2}{4} + \frac{3}{4} \right)(1 - \tilde{p}_1 - \tilde{p}_2) \right] \\ &= \frac{3}{16} + \frac{1}{4} \left( \tilde{p}_1 + \frac{1}{2}\tilde{p}_2 \right) \left( 1 - \tilde{p}_1 - \frac{1}{2}\tilde{p}_2 \right). \end{aligned}$$

Substituting this in (3.8) and again using (2.9), we obtain (4.1).

For the second part, we compute

$$\frac{1}{|\bar{E}'_n|} \sum_{(u,v) \in \bar{E}'_n} \bar{d}_u \bar{d}_v = \frac{2}{\ell_n} \sum_{v \in V_n} d_v^2,$$

and for  $p \geq 2$ ,

$$\frac{1}{|\bar{E}'_n|} \sum_{s \in \bar{V}_n} \bar{d}_s^p = \frac{1}{2\ell_n} 2^p (\ell_n/2) + \frac{1}{2\ell_n} \sum_{v \in V_n} d_v^p = 2^{p-2} + \frac{1}{2\ell_n} \sum_{v \in V_n} d_v^p,$$

As a result, when  $\mathbb{E}[D_{(n)}^3] \rightarrow \mathbb{E}[D^3] < \infty$ , we have

$$\rho(\bar{G}_n) \xrightarrow{\mathbb{P}} \frac{2\mu_2/\mu_1 - (1 + \mu_2/(2\mu_1))^2}{(2 + \mu_3/(2\mu_1)) - (1 + \mu_2/(2\mu_1))^2} < 0,$$

where  $\mu_p = \mathbb{E}[D^p]$ . □

### 4.3. Preferential Attachment Model

We discuss the general preferential attachment model (PAM), as formulated, for example, in [van der Hofstad 13, Chapter 8] or [Durrett 07, Chapter 4]. The PAM is a *dynamical* random graph model, and thus models a growing network. It is defined in terms of two parameters,  $m$ , which denotes the number of edges

of newly added vertices, and  $\delta > -m$ , which quantifies the tendency to attach to vertices that already have a high degree. We start by defining the model for  $m = 1$ .

We start with one vertex having one self-loop. Suppose we have the graph of size  $t$ , which we denote by  $G_t^{(1)}$ . Let  $i$  label the vertex that appeared at time  $i = 1, 2, \dots$ . Then,  $G_{t+1}^{(1)}$  is constructed by adding one extra vertex that has one edge, which forms a self-loop with probability  $(1 + \delta)/((2 + \delta)t + 1 + \delta)$  and, conditionally on  $G_t^{(1)}$ , attaches to a vertex  $v \in [t]$  with probability  $(D_i(t) + \delta)/((2 + \delta)t + 1 + \delta)$ , where  $D_i(t)$  is the random degree of vertex  $i$  in  $G_t^{(1)}$ . As a result, vertices with high degree have a higher probability to become attached, which explains the name *preferential attachment model*.

The model with  $m \geq 2$  is obtained from the model with  $m = 1$  as follows. Collapse vertices  $m(s - 1) + 1, \dots, ms$ , and all of their edges, in  $(G_t^{(1)})_{t \geq 1}$  with  $\delta$  replaced by  $\delta' = \delta/m$  to form vertex  $s$  in  $(G_t^{(m)})_{t \geq 1}$  with parameter  $\delta$ . It is well known (see e.g., [Bollobás et al. 01] where this was first derived for  $\delta = 0$  and [van der Hofstad 13, Theorem 8.3] as well as the references in [van der Hofstad 13] for a more detailed literature overview) that the resulting graph has an asymptotic degree sequence  $p_k$ , i.e.,

$$N_k(t)/t = \#\{i \in [t]: D_i(t) = k\}/t \xrightarrow{\mathbb{P}} p_k, \tag{4.7}$$

where, for  $k \geq m$ ,

$$p_k = (2 + \delta/m) \frac{\Gamma(k + \delta)\Gamma(m + 2 + \delta + \delta/m)}{\Gamma(m + \delta)\Gamma(k + 3 + \delta + \delta/m)}. \tag{4.8}$$

In particular, the PAM is scale free with power-law exponent  $\gamma = 2 + \delta/m$ . See [van der Hofstad 13, Section 8.2] for more details on the scale-free behavior of the PAM. The next theorem investigates the behavior of Pearson’s correlation coefficient as well as Spearman’s rho for the PAM:

**Theorem 4.4. (Convergence of degree-degree dependency measures for PAM).** *Let  $(G_t^{(m)})_{t \geq 1}$  be the PAM. Then*

$$\rho^{\text{rank}}(G_t^{(m)}) \xrightarrow{\mathbb{P}} \rho^{\text{rank}}, \tag{4.9}$$

while

$$\rho(G_t^{(m)}) \xrightarrow{\mathbb{P}} \begin{cases} 0 & \text{if } \delta \leq m, \\ \rho & \text{if } \delta > m, \end{cases} \tag{4.10}$$

where, abbreviating  $a = \delta/m$ ,

$$\rho = \frac{(m-1)(a-1)[2(1+m) + a(1+3m)]}{(1+m)[2(1+m) + a(5+7m) + a^2(1+7m)]}. \quad (4.11)$$

The value of  $\rho$  in (4.11) was predicted in [Dorogovtsev et al. 10], and we make this analysis mathematically rigorous. The remainder of the section is the proof of Theorem 4.4. It involves intermediate technical results formulated as Lemmas 4.5–4.9 below.

For the PAM, it will be convenient to direct the edges from young to old, so that there are  $mt$  directed edges. Let  $N_{k,l}(t)$  denote the number of directed edges  $e$  for which  $D_{\underline{e}}(t) = k$ ,  $D_{\overline{e}}(t) = l$ . We will prove that there exists a probability distribution  $(q_{k,l})_{k,l \geq m}$  such that

$$N_{k,l}(t)/(mt) \xrightarrow{\mathbb{P}} q_{k,l}. \quad (4.12)$$

Since a uniform directed edge oriented from young to old can be obtained by taking a uniform vertex and then a uniform edge coming out of this vertex, this proves (3.10) with

$$p_{kl} = \mathbb{P}(X = k, Y = l) = \frac{1}{2}(q_{k,l} + q_{l,k}). \quad (4.13)$$

In particular, by Theorem 3.2(a), this proves (4.9) in Theorem 4.4. We follow the proof of [van der Hofstad 13, Theorem 8.2], which, in turn, is strongly inspired by the proof in [Bollobás et al. 01].

Proofs for convergence of the degree sequence typically consist of two key steps. The first is a martingale concentration argument in Lemma 4.5.

**Lemma 4.5. (Convergence of degree-degree counts).** *For every  $k, l$ , there exists a  $C > 0$  such that,*

$$\mathbb{P}\left(\max_{k,l} |N_{kl}(t) - \mathbb{E}[N_{kl}(t)]| \geq C\sqrt{t \log t}\right) = o(1). \quad (4.14)$$

**Proof.** The proof for the degree distribution in [van der Hofstad 13] applies almost verbatim (see, in particular, [van der Hofstad 13, Proposition 8.4] and its proof). Indeed, the proof relies on a martingale argument. Define the Doob-martingale, for  $n = 0, \dots, t$ ,

$$M_n = \mathbb{E}[N_{kl}(t) \mid G_n^{(m)}].$$

The crucial observation is that  $(M_n)_{n=0}^t$  is a martingale with  $M_t = N_{kl}(t)$  and  $M_0 = \mathbb{E}[N_{kl}(t)]$  that satisfies

$$|M_n - M_{n-1}| \leq 4m. \quad (4.15)$$

We prove (4.15) below. The Azuma-Hoeffding inequality [Azuma 67, Hoeffding 63] then proves (4.14) for any  $C > 4[4m]^2$ . Indeed,

$$\mathbb{P}\left(|N_{kl}(t) - \mathbb{E}[N_{kl}(t)]| \geq A\right) = \mathbb{P}\left(|M_t - M_0| \geq A\right) \leq e^{-A^2/(2t[4m]^2)}.$$

Taking  $A = C\sqrt{t \log t}$  with  $C^2 > 4[4m]^2$  proves that

$$\mathbb{P}\left(|N_{kl}(t) - \mathbb{E}[N_{kl}(t)]| \geq C\sqrt{t \log t}\right) = o(1/t^2),$$

so that even

$$\begin{aligned} &\mathbb{P}\left(\max_{k,l} |N_{kl}(t) - \mathbb{E}[N_{kl}(t)]| \geq C\sqrt{t \log t}\right) \\ &\leq (mt)^2 \max_{k,l} \mathbb{P}\left(\max_{k,l} |N_{kl}(t) - \mathbb{E}[N_{kl}(t)]| \geq C\sqrt{t \log t}\right) = o(1). \end{aligned}$$

This completes the proof of Lemma 4.5, assuming (4.15).

We complete the proof by deriving (4.15). For this, it will be convenient to introduce some further notation. Let  $e \in [mt]$  label the edges. Let  $v_e = \lceil e/m \rceil$  denote the vertex from which the  $e$ th edge emanates, and  $V_e$  (which is a random variable) represent the vertex to which the  $e$ th edge points. Then,

$$N_{k,l}(t) = \sum_{e \in [mt]} \mathbb{1}_{\{D_{v_e}(t)=k, D_{V_e}(t)=l\}}.$$

As a result,

$$\begin{aligned} M_n - M_{n-1} &= \sum_{e \in [mt]} [\mathbb{P}(D_{v_e}(t) = k, D_{V_e}(t) = l \mid G_n) \\ &\quad - \mathbb{P}(D_{v_e}(t) = k, D_{V_e}(t) = l \mid G_{n-1})], \end{aligned}$$

where we abbreviate  $G_n = G_n^{(m)}$ . We let  $(G'_t)_{t \geq 0}$  denote the PAM with  $G'_{n-1} = G_{n-1}$ , while the evolution of  $(G'_t)_{t \geq 0}$  after time  $n - 1$  is the same in distribution as that of  $(G_t)_{t \geq 0}$ , but conditionally independent of it given  $G_{n-1} = G'_{n-1}$ . Let  $D'_i(t)$  denote the degree of vertex  $i$  in  $G'_t$ . Then,

$$\begin{aligned} \mathbb{P}(D_{v_e}(t) = k, D_{V_e}(t) = l \mid G_{n-1}) &= \mathbb{P}(D'_{v_e}(t) = k, D'_{V_e}(t) = l \mid G_{n-1}) \\ &= \mathbb{P}(D'_{v_e}(t) = k, D'_{V_e}(t) = l \mid G_{n-1}, G_n \setminus G_{n-1}), \end{aligned}$$

where  $G_n \setminus G_{n-1}$  is shorthand for the edges of  $G_n$  that are not in  $G_{n-1}$ . The last step is due to the conditional independence of the evolution after time  $n - 1$  in  $(G'_t)_{t \geq 0}$ . Thus,

$$\mathbb{P}(D_{v_e}(t) = k, D_{V_e}(t) = l \mid G_{n-1}) = \mathbb{P}(D'_{v_e}(t) = k, D'_{V_e}(t) = l \mid G_n).$$

We conclude that

$$M_n - M_{n-1} = \sum_{e \in [mt]} [\mathbb{P}(D_{v_e}(t) = k, D_{V_e}(t) = l \mid G_n) - \mathbb{P}(D'_{v_e}(t) = k, D'_{V_e}(t) = l \mid G_n)].$$

When  $V_e > n$ , clearly

$$\mathbb{P}(D_{v_e}(t) = k, D_{V_e}(t) = l \mid G_n) = \mathbb{P}(D'_{v_e}(t) = k, D'_{V_e}(t) = l \mid G_n),$$

as the degrees of vertices  $i$  with  $i > n$  are independent of  $G_n$ . Thus, we can restrict to  $V_e \leq n$ . Further, when  $v_e > n$ , then  $D_{v_e}(t)$  is independent of  $G_n$ , so that

$$\begin{aligned} & \mathbb{P}(D_{v_e}(t) = k, D_{V_e}(t) = l \mid G_n) - \mathbb{P}(D'_{v_e}(t) = k, D'_{V_e}(t) = l \mid G_n) \\ &= \mathbb{P}(D_{v_e}(t) = k) \left[ \mathbb{P}(D_{V_e}(t) = l \mid G_n) - \mathbb{P}(D'_{V_e}(t) = l \mid G_n) \right]. \end{aligned}$$

Note that  $D_{V_e}(n-1) = D'_{V_e}(n-1)$  a.s.,  $\mathbb{P}(D_{V_e}(t) = l \mid G_n, D_{V_e}(n) = j) = \mathbb{P}(D'_{V_e}(t) = l \mid D'_{V_e}(n) = j)$ , and

$$\begin{aligned} & \mathbb{P}(D'_{V_e}(t) = l \mid G_n, D'_{V_e}(n) = j) = \mathbb{P}(D'_{V_e}(t) = l \mid D'_{V_e}(n) = j) \\ &= \mathbb{P}(D_{V_e}(t) = l \mid D_{V_e}(n) = j). \end{aligned}$$

Thus, using that

$$\begin{aligned} & \mathbb{P}(D_{V_e}(t) = l \mid G_n) = \mathbb{E}[\mathbb{P}(D'_{V_e}(t) = l \mid D_{V_e}(n)) \mid G_n], \\ & \mathbb{P}(D'_{V_e}(t) = l \mid G_n) = \mathbb{E}[\mathbb{P}(D'_{V_e}(t) = l \mid D'_{V_e}(n)) \mid G_n], \end{aligned}$$

we obtain at

$$|\mathbb{P}(D'_{V_e}(t) = l \mid D_{V_e}(n)) - \mathbb{P}(D'_{V_e}(t) = l \mid D'_{V_e}(n))| \leq \mathbb{1}_{\{D_{V_e}(n) \neq D'_{V_e}(n)\}}.$$

Taking expectations yields

$$\begin{aligned} & |\mathbb{P}(D_{v_e}(t) = k, D_{V_e}(t) = l \mid G_n) - \mathbb{P}(D'_{v_e}(t) = k, D'_{V_e}(t) = l \mid G_n)| \\ & \leq \mathbb{P}(D_{V_e}(n) \neq D'_{V_e}(n) \mid G_n). \end{aligned}$$

In a similar way, we see that for  $v_e \leq n$ ,

$$\begin{aligned} & |\mathbb{P}(D_{v_e}(t) = k, D_{V_e}(t) = l \mid G_n) - \mathbb{P}(D'_{v_e}(t) = k, D'_{V_e}(t) = l \mid G_n)| \\ & \leq \mathbb{P}(D_{V_e}(n) \neq D'_{V_e}(n) \mid G_n) + \mathbb{P}(D_{v_e}(n) \neq D'_{v_e}(n) \mid G_n). \end{aligned}$$

We conclude that

$$\begin{aligned} |M_n - M_{n-1}| & \leq \sum_{e \in [mt]} [\mathbb{P}(D_{V_e}(n) \neq D'_{V_e}(n) \mid G_n) + \mathbb{P}(D_{v_e}(n) \neq D'_{v_e}(n) \mid G_n)] \\ & \leq 4m. \end{aligned}$$

□

We continue with the proof of (4.12). The second key step of the proof of (4.12) is to prove that, for each  $k, l$ ,

$$\lim_{t \rightarrow \infty} \mathbb{E} [N_{kl}(t)] / (mt) = q_{k,l}. \tag{4.16}$$

We sum over the vertex  $s$  that has degree  $l$  at time  $t$ , and condition on the degree  $r \geq m$  of the vertex to which the edge of vertex  $s$  is attached. This yields

$$\begin{aligned} \mathbb{E} [N_{kl}(t)] &= m \sum_{s=1}^t \sum_{r \geq m} \frac{(r + \delta)}{(2m + \delta)s} \mathbb{E} [N_r(s)] \\ &\quad \times [\mathbb{P} (B_{r+1}[s + 1, t] = k, B_m[s + 1, t] = l) + O(1/s)], \end{aligned} \tag{4.17}$$

where  $B_m[s + 1, t]$  is  $m$  plus the number of edges attached to vertex  $s$  between time  $s + 1$  and  $t$ , while  $B_{r+1}[s + 1, t]$  is  $r$  plus the number of further edges attached to the vertex of degree  $r$  to which the edge of vertex  $s$  is attached. The  $O(1/s)$  term is due to contributions where at least *two* edges of vertex  $s$  are attached to the same vertex of degree  $r$ , and also due to the fact that the probability of attaching the  $j$ th edge of vertex  $s$  to a vertex of degree  $r$  at time  $s$  is actually equal to  $\frac{(r+\delta)}{(2m+\delta)s+(j-1)(2+\delta/m)+1+\delta/m}$ , which is  $\frac{(r+\delta)}{(2m+\delta)s}(1 + O(1/s))$ . Further,

$$\begin{aligned} &\mathbb{P} (B_{r+1}[s + 1, t] = k, B_m[s + 1, t] = l) \\ &= \mathbb{P} (B_{r+1}[s + 1, t] = k) \mathbb{P} (B_m[s + 1, t] = l) + O(1/t), \end{aligned}$$

since the dependence between the two probabilities is entirely due to the fact that edges that contribute to  $B_{r+1}[s + 1, t]$  cannot contribute to  $B_m[s + 1, t]$ . Indeed,  $(B_{r+1}[s + 1, t], B_m[s + 1, t])$  is equal in distribution to the number of balls in two urns at time  $m(t - s)$ , where we start with  $r + 1$  and  $m$  balls at time 0, and in each draw, we draw a ball in each of the urns with probability equal to the number of balls plus  $\delta$  and then replace it with two balls. Knowing how many balls are put into the first urn gives us information only about how many balls cannot be put into the second urn, so the balls in the different urns are close to independent. We study these probabilities now:

**Lemma 4.6. (Growth of degrees in PAM).** *For all  $k \geq r \geq m$  and  $a \in (0, 1)$ ,*

$$\lim_{s \rightarrow \infty} \mathbb{P} (B_r[as, s] = k) = P_k(a; r),$$

where, for each  $r \geq m$  and  $a \in (0, 1)$ ,  $(P_k(a; r))_{k \geq r}$  is a probability measure.

**Proof.** We note that  $(B_r[s, ts])_{t \geq 1} \xrightarrow{d} (Z_t)_{t \geq 1}$ , as  $s \rightarrow \infty$ , where  $(Z_t)_{t \geq 0}$  is a pure birth process, which increases by 1 at rate  $m(Z_t + \delta) / ((2m + \delta)t)$  at time  $t$ . Indeed, when  $B_r[s, ts] = k$ , then each of the  $m$  edges of vertex  $st + 1$  has

probability  $(k + \delta)/[(2m + \delta)(st)] + O(1/s^2)$  of being attached to the vertex that has degree  $k$  at time  $ts$ , and thus of increasing  $B_r[s, ts]$  to  $k + 1$ . Thus, within a short time interval  $[t, t + dt]$  and conditionally on  $B_r[s, ts] = k$ , the probability that  $B_r[s, (t + dt)s] = k + 1$  is equal to

$$sdt \left[ m(k + \delta)/[(2m + \delta)(st)] + O(1/s^2) + o(1) \right] \rightarrow dt \frac{m(k + \delta)}{(2m + \delta)t} + o(dt),$$

as  $s \rightarrow \infty$ . This is the birth rate of the pure birth process  $(Z_t)_{t \geq 1}$ .

We next study the limiting birth process, for which is it useful to make a time change. With  $b_t = Z_{e^{(2+\delta/m)t}}$ ,  $(b_t)_{t \geq 0}$  is a birth process that grows at rate  $b_t + \delta$  at time  $t$ . Define

$$f_{r,k}(t) = \mathbb{P}(b_t = k \mid b_0 = r).$$

Then,

$$\frac{\partial}{\partial t} f_{r,k}(t) = -(k + \delta)f_{r,k}(t) + (k - 1 + \delta)f_{r,k-1}(t).$$

This set of differential equations is solved by  $f_{r,r}(t) = e^{-(r+\delta)t}$  and, for  $k \geq r + 1$ ,

$$f_{r,k}(t) = (k - 1 + \delta)e^{-(k+\delta)t} \int_0^t e^{(k+\delta)s} f_{r,k-1}(s) ds.$$

This can be solved, for  $k \geq r + 1$ , by

$$f_{r,k}(t) = \mathbb{P}(b_t = i \mid b_0 = r) = \frac{\Gamma(k + \delta)}{\Gamma(r + \delta)} e^{-(k+\delta)t} \sum_{j=0}^{k-r} \alpha_{j,k} e^{jt},$$

where  $\alpha_{0,k} = -\sum_{j=0}^{k-1} \alpha_{j,k-1}/(j + 1)$ , while, for  $j \geq 1$ ,

$$\alpha_{j,k} = \alpha_{j-1,k-1}/j.$$

As a result, for all  $a \in (0, 1)$ ,

$$\lim_{t \rightarrow \infty} \mathbb{P}(B_r[at, t] = k) = \mathbb{P}(Z_{1/a} = k \mid Z_1 = r) = f_{r,k}((2 + \delta/m)^{-1} \log(1/a)).$$

Note that  $P_r(a; r)$  is the probability that the birth process has no births. We thus compute that  $P_r(a; r) = f_{r,r}((2 + \delta/m)^{-1} \log(1/a)) = a^{(r+\delta)/(2+\delta/m)}$  for  $k = r$ , while

$$\begin{aligned} P_k(a; r) &= f_{r,k}((2 + \delta/m)^{-1} \log(1/a)) \\ &= \frac{\Gamma(k + \delta)}{\Gamma(r + \delta)} a^{(k+\delta)/(2+\delta/m)} \sum_{j=0}^{k-r} \alpha_{j,k} a^{-j/(2+\delta/m)}. \end{aligned}$$

□



We continue from (4.17), and rewrite it as

$$\begin{aligned} \mathbb{E} [N_{kl}(t)]/(mt) &= \sum_{r \geq m} \mathbb{E} \left[ \frac{(r + \delta)}{(2m + \delta)Ut} \mathbb{E} [N_r(Ut)] \mathbb{P} (B_{r+1}[Ut, t] = k \mid U) \right. \\ &\quad \left. \times \mathbb{P} (B_m [Ut, t] = l \mid U) \right] + O(\log t/t), \end{aligned} \tag{4.18}$$

where  $U$  has a uniform distribution, we interpret  $Ut = \lceil Ut \rceil$ , and the outer expectation is over  $U$  only. Using that  $\mathbb{E} [N_r(s)]/s = p_r + O(1/s)$  (see [van der Hofstad 13, Proposition 8.4]), we further arrive at

$$\begin{aligned} \mathbb{E} [N_{kl}(t)]/(mt) &= \sum_{r \geq m} \frac{r + \delta}{2m + \delta} p_r \mathbb{E} [\mathbb{P} (B_{r+1}[Ut, t] = k \mid U) \mathbb{P} (B_m [Ut, t] = l \mid U)] \\ &\quad + o(1). \end{aligned} \tag{4.19}$$

By Lemma 4.6, this converges to

$$\mathbb{E} [N_{kl}(t)]/(mt) \rightarrow q_{k,l} \equiv \sum_{r \geq m} \frac{r + \delta}{2m + \delta} p_r \mathbb{E} [P_k(U; r) P_l(U; m)]. \tag{4.20}$$

This proves (4.16), and thus, by Theorem 3.2(a), proves the convergence of the rank correlation in (4.9) in Theorem 4.4.

For the convergence of the correlation coefficient in (4.10) in Theorem 4.4, we aim to use Theorem 3.2(b) and thus start by investigating the convergence of moments of  $X_n$ . By (3.3), and letting  $\mathbb{E}_n$  denote the conditional expectation given  $G_n$ ,

$$\mathbb{E}_n [X_n^2] = \frac{1}{n} \sum_{i \in [n]} D_i(n)^3.$$

Thus, we are led to studying sums of powers of degrees. To analyze the limit of sums of powers of degrees, we rely on the following lemma:

**Lemma 4.7. (Sum of powers of degrees in PAM).** *For all  $p < \gamma = 2 + \delta/m$ ,*

$$\frac{1}{n} \sum_{i \in [n]} D_i(n)^p \xrightarrow{\mathbb{P}} \mu_p = \sum_{k \geq m} k^p p_k < \infty.$$

**Proof.** We note that  $\sum_{i \in [n]} D_i(n)^p = \sum_{k \geq m} k^p N_k(n)$ . Under the conditions stated, for every  $k_n \rightarrow \infty$ ,

$$\sum_{k \geq m} k^p N_k(n) = \sum_{m \leq k \leq k_n} k^p N_k(n) + o_{\mathbb{P}}(n).$$

This follows since, for any  $\varepsilon > 0$ ,  $k > k_n$  implies that  $k^\varepsilon/k_n^\varepsilon > 1$ , so that

$$\sum_{k > k_n} k^p N_k(n) \leq k_n^{-\varepsilon} \sum_{k \geq m} k^{p+\varepsilon} N_k(n) = k_n^{-\varepsilon} \frac{1}{n} \sum_{i \in [n]} D_i(n)^{p+\varepsilon}.$$

By the analysis in [van der Hofstad 13, Sections 8.1 and 8.6], when  $p + \varepsilon < \gamma + 1 = 3 + \delta/m$ ,

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left[ \frac{1}{n} \sum_{i \in [n]} D_i(n)^{p+\varepsilon} \right] < \infty.$$

Therefore, by the Markov inequality,  $\sum_{k > k_n} k^p N_k(n) = o_{\mathbb{P}}(n)$ .

Now, since  $\max_k |N_k(n) - p_k| \leq \sqrt{Cn \log n}$  whp by [van der Hofstad 13, Proposition 8.4],

$$\sum_{m \leq k \leq k_n} k^p N_k(n) = t \sum_{m \leq k \leq k_n} k^p p_k + O_{\mathbb{P}}(k_n^{p+1} \sqrt{n \log n}).$$

This proves the claim. □

It follows from Lemma 4.7 that, for  $3 < \gamma = 2 + \delta/m$ ,

$$\mathbb{E}_n[X_n^2] = \frac{1}{n} \sum_{i \in [n]} D_i(n)^3 = B(1 + o_{\mathbb{P}}(1)),$$

where  $B$  is a constant. As a result,

$$\rho(G_n) \xrightarrow{a.s.} \rho = \text{Cov}(X, Y) / \text{Var}(X) = \frac{\sum_{k,l} k l q_{k,l} - \mathbb{E}[X]^2}{\mathbb{E}[X^2] - \mathbb{E}[X]^2}. \tag{4.21}$$

This proves (4.10) in Theorem 4.4 when  $\delta > m$ . For  $\gamma < 3$ , instead,  $D_1(n)/n^{1/\gamma} \xrightarrow{a.s.} \xi$ , for some strictly positive random variable  $\xi$  (see e.g., [van der Hofstad 13, Sections 8.1 and 8.6]). Therefore,  $\mathbb{E}_n[X_n^2] \geq \xi^3 n^{3/\gamma-1}(1 + o(1))$ . Further, the majority of edges of high-degree vertices is young, so that

$$\mathbb{E}_n[X_n Y_n] = o_{\mathbb{P}}(n^{3/\gamma-1}). \tag{4.22}$$

Indeed, fix  $T_n$  such that  $T_n \rightarrow \infty$  and  $T_n = o(n)$ . There are, at most,  $mT_n$  edges between vertices with index, at most,  $T_n$ , and, since the maximal degree is  $O_{\mathbb{P}}(n^{1/\gamma})$ , these contribute, at most,  $O_{\mathbb{P}}(n^{2/\gamma-1}T_n)$ . For the other edges, one of the vertices involved was born after time  $T_n$ . Since  $\max_{i \geq T_n} D_i(n) = o_{\mathbb{P}}(n^{1/\gamma})$ , the contribution of these edges is, at most,

$$o_{\mathbb{P}}(n^{1/\gamma}) \mathbb{E}_n[X_n + Y_n].$$

In turn,  $\mathbb{E}_n[X_n + Y_n] = O_{\mathbb{P}}(n^{(2/\gamma-1) \wedge 1})$ , which completes the proof of (4.22). This implies that  $\rho(G_n) \xrightarrow{\mathbb{P}} 0$ , which proves (4.10) in Theorem 4.4 when  $\delta < m$ .

For  $\delta = m$ , so that  $\gamma = 3$ ,  $\sum_{i \in [n]} D_i(n)^3 = \Theta_{\mathbb{P}}(n \log n)(1 + o_{\mathbb{P}}(1))$ . As a result, also in this case  $\rho(G_n) \xrightarrow{a.s.} 0$  for  $\delta \leq m$ .  $\square$

We continue with the proof of (4.11) in Theorem 4.4. To compute expectations involving  $X$ , we often rely on the following lemma:

**Lemma 4.8. (Degree on one side of uniform edge).** *For every function  $f: \mathbb{N} \rightarrow \mathbb{R}$ ,*

$$\mathbb{E}[f(X)] = \sum_{k \geq m} f(k) \frac{k p_k}{2m}.$$

**Proof.** Let  $f$  be bounded, and let  $X_n$  be the degree at the bottom of a uniform edge. Then,

$$\begin{aligned} \mathbb{E}[f(X_n) \mid G_n^{(m)}] &= \frac{1}{|E'_n|} \sum_{e \in E'_n} f(D_{\underline{e}}(n)) = \frac{1}{2mn} \sum_{v \in [n]} f(D_v(n)) D_v(n) \\ &= \frac{1}{2m} \sum_{k \geq m} f(k) k N_k(n) / n. \end{aligned}$$

Taking the limit of  $n \rightarrow \infty$ , and using that  $N_k(n)/n \xrightarrow{\mathbb{P}} p_k$ , as well as  $X_n \xrightarrow{d} X$ , proves the claim.  $\square$

Lemma 4.8 allows us to identify the r.h.s. of (4.21) as

$$\rho = \text{Cov}(X, Y) / \text{Var}(X) = \frac{(2m)^2 \sum_{k,l} k l q_{k,l} - \lambda_2^2}{2m \lambda_3 - \lambda_2^2},$$

where  $\lambda_a = \sum_{k \geq m} k^a p_k$ . To identify the limit, we follow [Dorogovtsev et al. 10]. Recall the definition of  $p_{kl}$  in (4.13).

**Lemma 4.9. (Asymptotic degree-degree distribution for PAM).** *For all  $k, l \geq m$ ,*

$$\begin{aligned} p_{kl} &= \mathbb{P}(X = k, Y = l) \tag{4.23} \\ &= (2 + \delta/m) \frac{\Gamma(m + 2 + \delta + \delta/m)}{\Gamma(m + \delta)^2} \frac{\Gamma(l + \delta) \Gamma(k + \delta)}{\Gamma(k + 2 + \delta) \Gamma(l + k + 2 + 2\delta + \delta/m)} \\ &\quad \times \left[ \sum_{j=m+1}^k \binom{k+l-j-m}{l-m} \binom{j+k+2+2\delta+\delta/m}{k+1+\delta} \right. \\ &\quad \left. + \sum_{j=m+1}^l \binom{k+l-j-m}{k-m} \binom{j+l+2+2\delta+\delta/m}{l+1+\delta} \right]. \end{aligned}$$

Consequently, (4.11) follows.

**Proof.** To compute  $\mathbb{P}(X = k, Y = l)$ , we let  $M_{kl}(t)$  denote the number of edges at time  $t$  where one side has degree  $k$  and the other side degree  $l$ , so that

$$p_{kl} = \lim_{t \rightarrow \infty} \mathbb{E}[M_{kl}(t)]/(2mt).$$

We note that  $M_{kl}(t)$  satisfies the recursion relation

$$\begin{aligned} & \mathbb{E}[M_{kl}(t+1)] - \mathbb{E}[M_{kl}(t)] \\ &= m \frac{(k \vee l) - 1 + \delta}{(2m + \delta)t} \mathbb{E}[N_{k \vee l - 1}(t)] \mathbb{1}_{\{k \wedge l = m\}} \\ & \quad + m \frac{k - 1 + \delta}{(2m + \delta)t} \mathbb{E}[M_{k-1,l}(t)] + m \frac{l - 1 + \delta}{(2m + \delta)t} \mathbb{E}[M_{k,l-1}(t)] \\ & \quad - m \frac{k + \delta}{(2m + \delta)t} \mathbb{E}[M_{k,l}(t)] - m \frac{l + \delta}{(2m + \delta)t} \mathbb{E}[M_{k,l}(t)] + O(1/t^2). \end{aligned}$$

It is not clear that the left-hand side converges because we know only that  $\mathbb{E}[M_{k,l}(t)]/(2mt) \rightarrow p_{kl}$ , and we will show this now. Indeed, since  $\mathbb{E}[M_{k,l}(t)]/(2mt) \rightarrow p_{kl}$  and  $\mathbb{E}[N_k(t)]/t \rightarrow p_k$ , we arrive at the claim that, for all  $k, l$  with  $k \vee l \geq m + 1$ ,

$$\begin{aligned} & \lim_{t \rightarrow \infty} \mathbb{E}[M_{kl}(t+1)] - \mathbb{E}[M_{kl}(t)] \\ &= 2m^2 \frac{(k \vee l) - 1 + \delta}{2m + \delta} p_{k-1} \mathbb{1}_{\{k \wedge l = m\}} + 2m^2 \frac{k - 1 + \delta}{2m + \delta} p_{k-1,l} \\ & \quad + 2m^2 \frac{l - 1 + \delta}{2m + \delta} p_{k,l-1} - 2m^2 \frac{k + l + 2\delta}{2m + \delta} p_{k,l}. \end{aligned}$$

Since  $\lim_{t \rightarrow \infty} \mathbb{E}[M_{kl}(t)]/(2mt) = p_{kl}$ , we must therefore have that  $\lim_{t \rightarrow \infty} \mathbb{E}[M_{kl}(t+1)] - \mathbb{E}[M_{kl}(t)] = 2mp_{kl}$ , so that

$$\begin{aligned} p_{kl} &= m \frac{(k \vee l) - 1 + \delta}{2m + \delta} p_{k \vee l - 1} \mathbb{1}_{\{k \wedge l = m\}} + m \frac{k - 1 + \delta}{2m + \delta} p_{k-1,l} + m \frac{l - 1 + \delta}{2m + \delta} p_{k,l-1} \\ & \quad - m \frac{k + l + 2\delta}{2m + \delta} p_{k,l}, \end{aligned}$$

and

$$\begin{aligned} (k + l + 2 + 2\delta + \delta/m)p_{kl} &= ((k \vee l) - 1 + \delta)p_{k \vee l - 1} \mathbb{1}_{\{k \wedge l = m\}} \\ & \quad + (k - 1 + \delta)p_{k-1,l} + (l - 1 + \delta)p_{k,l-1}. \quad (4.24) \end{aligned}$$

This is equivalent to [Dorogovtsev et al. 10, (12)]. This can be worked out to yield

$$p_{kl} = \sum_{j=m+1}^k \binom{k+l-j-m}{k-j} \frac{\Gamma(k+\delta)}{\Gamma(j-1+\delta)} \frac{\Gamma(l+\delta)}{\Gamma(m+\delta)} \frac{\Gamma(j+k+2+2\delta+\delta/m)}{\Gamma(l+k+3+2\delta+\delta/m)} p_{j-1} \\ + \sum_{j=m+1}^l \binom{k+l-j-m}{l-j} \frac{\Gamma(k+\delta)}{\Gamma(j-1+\delta)} \frac{\Gamma(l+\delta)}{\Gamma(m+\delta)} \frac{\Gamma(j+l+2+2\delta+\delta/m)}{\Gamma(l+k+3+2\delta+\delta/m)} p_{j-1}.$$

Substituting (4.8), we arrive at (4.23).

The computation to go from (4.24) to (4.11) is performed in [Dorogovtsev et al. 10, (12)], and applies verbatim.  $\square$

#### 4.4. Asymptotically Random Pearson’s Coefficient: Collection of Complete Bipartite Graphs

In this section, we present an example where  $\rho(G_n)$  in (3.4) converges to a random variable when the number of vertices tends to infinity. For  $|V_n| = n$ , under the assumptions of Theorem 3.1, we have

$$\sum_{(u,v) \in E'_n} D_u D_v \leq \max_{v \in V_n} d_v \sum_{(u,v) \in E'_n} D_u = \max_{v \in V_n} D_v \left( \sum_{v \in V_n} D_v^2 \right) \leq C^2 n^{1/\gamma + (2/\gamma \vee 1)}, \tag{4.25}$$

$$\sum_{(u,v) \in E'_n} D_u D_v \geq \max_{v \in V_n} D_v \geq cn^{1/\gamma}, \tag{4.26}$$

$$\sum_{(u,v) \in E'_n} D_u D_v \geq \sum_{v \in V_n} D_v^2 \geq cn^{2/\gamma \vee 1}. \tag{4.27}$$

Further, from the proof of Theorem 3.1, we know that

$$\sum_{v \in V_n} D_v^3 \geq (\max_{v \in V_n} D_v)^3 \geq c^3 n^{3/\gamma}, \tag{4.28}$$

and

$$\frac{1}{|E'_n|} \left( \sum_{v \in V_n} D_v^2 \right)^2 \leq (C^2/c) n^{(4/\gamma - 1) \vee 1}, \tag{4.29}$$

where we see that (4.29) is vanishing compared to (4.28). The convergence of (3.4) to a random variable can take place only if the crossproducts on the left-hand side of (4.25–4.27) are of the same order of magnitude as the left-hand side of (4.28). As we see from the above, this is possible for  $\gamma \in (1, 3)$ .

Below we present an example where  $\rho(G_n)$  indeed converges to a random variable. However, due to slow convergence, a substantially larger computational capacity is needed in order to approximate the limiting distribution.

Take  $((X_i, Y_i))_{i=1}^n$  to be an i.i.d. sample of integer random variables as in (2.10), where  $\alpha_1 = \alpha_2 = \beta_1 = b$ ,  $\beta_2 = ab$  for some  $b > 0$  and  $a > 1$ . Then, for  $i = 1, \dots, n$ , we create a complete bipartite graph of  $X_i$  and  $Y_i$  vertices, respectively. These  $n$  complete bipartite graphs are not connected to one another. We denote such a collection of  $n$  bipartite graphs by  $G_n$ . The graph  $G_n$  has  $|V_n| = \sum_{i=1}^n (X_i + Y_i)$  vertices and  $|E'_n| = 2 \sum_{i=1}^n X_i Y_i$  directed edges. Further, if  $D_v$  denotes the random degree of vertex  $v$ , then we obtain

$$\sum_{v \in V_n} D_v^p = \sum_{i=1}^n (X_i^p Y_i + Y_i^p X_i), \quad \sum_{(u,v) \in E'_n} D_u D_v = 2 \sum_{i=1}^n (X_i Y_i)^2.$$

Assume that the  $\xi_j$ 's in (2.10) satisfy (2.11) with  $\gamma \in (2, 4)$ , so that  $\mathbb{E}[\xi^2] < \infty$ , but  $\mathbb{E}[\xi^4] = \infty$ . As a result,  $|E'_n|/n \xrightarrow{P} 2\mathbb{E}[XY] < \infty$  and  $\frac{1}{n} \sum_{v \in V} D_v^2 \xrightarrow{P} \mathbb{E}[XY(X+Y)] < \infty$  when  $\gamma \in (3, 4)$ , while, for  $\gamma \in (2, 3)$ ,

$$n^{-3/\gamma} \sum_{v \in V} D_v^2 = n^{-3/\gamma} \sum_{i=1}^n (X_i^2 Y_i + Y_i^2 X_i) \xrightarrow{d} Z, \tag{4.30}$$

for some random variable  $Z$ . (For  $\gamma = 3$ , this sum grows as a slowly varying function in  $n$ , but this case is very similar and will thus be omitted). Further,

$$\begin{aligned} n^{-4/\gamma} b^{-4} \sum_{i=1}^n (X_i^3 Y_i + Y_i^3 X_i) &\xrightarrow{d} (a^3 + a)Z_1 + 2Z_2, \\ n^{-4/\gamma} b^{-4} \sum_{i=1}^n (X_i Y_i)^2 &\xrightarrow{d} a^2 Z_1 + Z_2, \end{aligned}$$

where  $Z_1$  and  $Z_2$  are two independent stable distributions with parameter  $\gamma/4$ . Therefore, using (3.4) and the fact that  $4/\gamma > (6/\gamma - 1) \wedge 1$ , we arrive at

$$\rho(G_n) \xrightarrow{d} \frac{2a^2 Z_1 + 2Z_2}{(a + a^3)Z_1 + 2Z_2}, \text{ as } n \rightarrow \infty,$$

which is a proper random variable taking values in  $(2a/(1 + a^2), 1)$ .

For convergence of the rank correlation, we note that

$$\mathbb{P}(X_n = k, Y_n = l) \rightarrow \mathbb{P}(X = k, Y = l) = \frac{kl}{\mathbb{E}[X_1 Y_1]} \mathbb{P}(X_1 = k, Y_1 = l),$$

where we recall that  $(X_1, Y_1)$  is as in (2.10), whereas  $(X, Y)$  are the degrees at either side of a uniformly chosen edge. Thus, convergence of the rank correlation follows from Theorem 3.2(a).

## 5. Numerical Results

In this section, we present numerical examples that illustrate our results.

### 5.1. Numerical Results for Configuration Models and Preferential Attachment Model

We have generated random graphs of different sizes using the configuration model in Section 4.1, the configuration model with intermediate vertices in Section 4.2, and the preferential attachment model (PAM) in Section 4.3. For the undirected preferential attachment model, we use the basic version with  $m = 1$  and  $\delta = 0$ , which implies  $\gamma = 2$ . In both configuration models (without and with intermediate vertices) we generate the degree sequences by rounding up i.i.d. values of a continuous random variable  $\eta$  with Pareto distribution:  $\mathbb{P}(\eta > x) = 4x^{-2}$ ,  $x > 2$ . The exponent  $\gamma = 2$  is chosen for a fair comparison to PAM, and all degrees are at least three for the strongest disassortativity in the model with intermediate vertices, see (4.1). In the case of the configuration graph in Section 4.1, we consider two versions: the original model with self-loops and double edges present, and the model where self-loops and double edges are removed. The rank correlation coefficient  $\rho^{\text{rank}}(G)$  is computed as in (3.2). The results are presented in Table 3.

The results for the configuration model with intermediate vertices confirm our findings in Section 4.2: Pearson's coefficient converges to zero, whereas Spearman's rho quickly converges to  $-0.75$ , revealing the strong negative dependence. For the PAM, Pearson's coefficient converges to zero, as indicated in Theorem 3.1, and Spearman's rank correlation clearly indicates a negative dependence. This can be understood by noting that the majority of edges of vertices with high degrees, that are old vertices, come from vertices that are added late in the graph growth process and thus have small degrees. However, by the growth mechanism of the PAM, vertices with low degrees are more likely to be connected to vertices having high degrees, which indeed suggests negative degree-degree dependencies.

We emphasize that, under given model assumptions, the graphs of different sizes have been constructed by the same algorithm. Thus, their mixing patterns are exactly the same. As we predicted, the Pearson correlation coefficient fails to reflect the intrinsic properties of the model because its absolute value decreases with the graph size, and converges to zero for all models. On the contrary, Spearman's rho consistently shows neutral mixing for the classical configuration model, moderately disassortative mixing for the preferential attachment graph, and strongly disassortative mixing for the configuration model with intermediate vertices.

Model	Characteristic	$n$			
		$10^2$	$10^3$	$10^4$	$10^5$
Configuration model with self-loops and double edges	$\mathbb{E}_N(\rho(G_n))$	-0.0070	-0.0018	-0.0011	0.0006
	$\sigma_N(\rho(G_n))$	0.0735	0.0221	0.0077	0.0017
	$\mathbb{E}_N(\rho^{\text{rank}}(G_n))$	0.0056	-0.0098	-0.0036	0.0005
	$\sigma_N(\rho^{\text{rank}}(G_n))$	0.0504	0.0150	0.0046	0.0019
Configuration model without self-loops and double edges	$\mathbb{E}_N(\rho(G_n))$	-0.0713	-0.0226	-0.0150	-0.0032
	$\sigma_N(\rho(G_n))$	0.0546	0.0188	0.0092	0.0029
	$\mathbb{E}_N(\rho^{\text{rank}}(G_n))$	-0.0409	-0.0094	-0.0032	-0.0006
	$\sigma_N(\rho^{\text{rank}}(G_n))$	0.0700	0.0201	0.0083	0.0021
Configuration model with intermediate vertices	$\mathbb{E}_N(\rho(\bar{G}_n))$	-0.2804	-0.1346	-0.0572	-0.0291
	$\sigma_N(\rho(\bar{G}_n))$	0.0742	0.0517	0.0279	0.0147
	$\mathbb{E}_N(\rho^{\text{rank}}(\bar{G}_n))$	-0.7523	-0.7498	-0.7498	-0.7500
	$\sigma_N(\rho^{\text{rank}}(\bar{G}_n))$	0.0081	0.0025	0.0008	0.0003
Preferential attachment	$\mathbb{E}_N(\rho(G_n))$	-0.2682	-0.1282	-0.0608	-0.0272
	$\sigma_N(\rho(G_n))$	0.0575	0.0271	0.0132	0.0064
	$\mathbb{E}_N(\rho^{\text{rank}}(G_n))$	-0.4347	-0.4263	-0.4288	-0.4289
	$\sigma_N(\rho^{\text{rank}}(G_n))$	0.0627	0.0272	0.0065	0.0020

**Table 3.** Estimated mean and standard deviation of  $\rho(G_n)$  and  $\rho^{\text{rank}}(G_n)$  obtained from 20 realizations of  $G_n$  for random graph models in Sections 4.1–4.3.

## 5.2. Numerical Results for Collections of Bipartite Graphs

We next compute the degree-degree dependencies in the collection of bipartite graphs discussed in Section 4.4. In Table 4 we present numerical results for  $\rho(G_n)$  and  $\rho^{\text{rank}}(G_n)$ . Here we choose  $b = 1/2$ ,  $a = 2$ ,  $\xi$  has a generalized Pareto distribution  $\mathbb{P}(\xi > x) = (1 + (x - 1)/2.8)^{-2.8}$ ,  $x > 1$ , and the degrees  $X$  and  $Y$  are obtained by rounding up the values in (2.10).

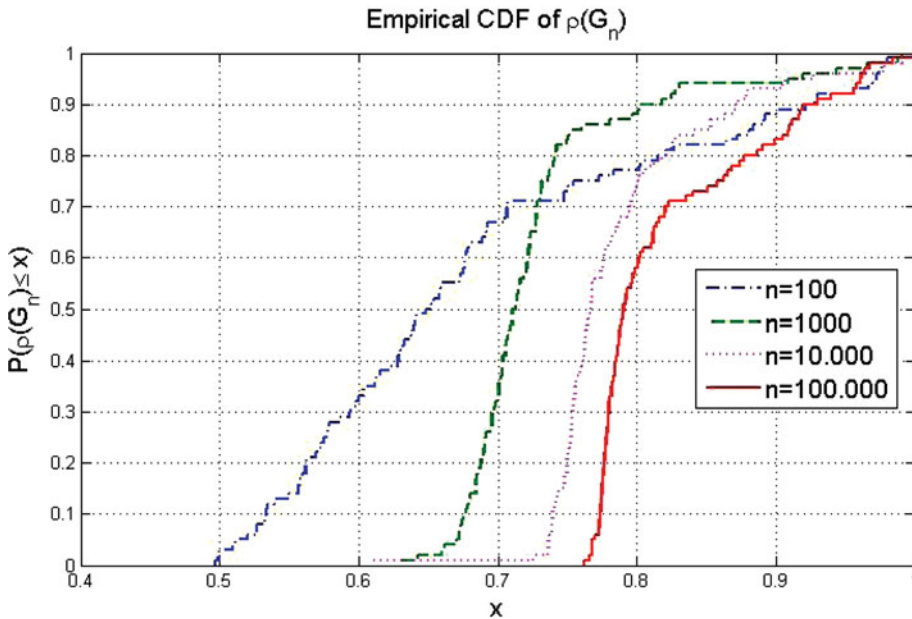
Note that in this model there is a genuine dependence between the correlation measure and the graph size. Indeed, if  $n = 1$  then the assortativity coefficient equals  $-1$  because nodes with larger degrees are connected to nodes with smaller degrees. However, when the graph size grows, the positive correlations start dominating because of the positive linear dependence between  $X$  and  $Y$ . We see that, again, the rank correlation captures the relation faster and gives consistent results with decreasing dispersion of values. Finally, Figure 2 shows the changes in the empirical distribution of  $\rho(G_n)$  as  $n$  grows. It is clear that a part of the



$n$	$10^2$	$10^3$	$10^4$	$10^5$
$\mathbb{E}_N(\rho(G_n))$	0.6554	0.7247	0.8042	0.8265
$\sigma_N(\rho(G_n))$	0.1145	0.1406	0.0689	0.0654
$\mathbb{E}_N(\rho^{\text{rank}}(G_n))$	0.7575	0.7950	0.8526	0.8615
$\sigma_N(\rho^{\text{rank}}(G_n))$	0.0735	0.1377	0.0218	0.0074

**Table 4.** Estimated mean and standard deviation of  $\rho(G_n)$  and  $\rho^{\text{rank}}(G_n)$  for the collection of  $n$  complete bipartite graphs. The number of realizations for each graph size is 20.

probability mass is spread over the interval  $(0.8, 1)$ . In the limit,  $\rho(G_n)$  has a nonzero density on this interval. The difference between the crossproducts and the expectation squared in  $\rho(G_n)$  is only of the order  $n^{1-2/\gamma}$ , which is about  $n^{0.29}$  in our example, thus, the convergence is too slow to be observed at  $n = 100,000$ .



**Figure 2.** The empirical distribution function  $\mathbb{P}(\rho(G_n) \leq x)$  for 100 observed values of  $\rho(G_n)$ , where  $G_n$  is a collection of  $n$  complete bipartite graphs.

nr	Dataset	Description	# nodes	# edges	max degree	$\rho(G_n)$	$\rho^{\text{rank}}(G_n)$	$\rho^-(G_n)$
(i)	stanford-cs	web domain	9,914	54,854	340	-.1656	-.1627	-.4648
(ii)	eu-2005	.eu web domain	862,664	5,477,938	68,963	-.0562	-.2525	-.0670
(iii)	uk@100,000	.uk web crawl	100,000	5,559,150	55,252	-.6536	-.5676	-1.117
(iv)	uk@1,000,000	.uk web crawl	1,000,000	77,123,940	403,441	-.0831	-.5620	-.0854
(v)	euron	e-mail exchange	69,244	506,898	1,634	-.1599	-.6827	-.1932
(vi)	dblp-2010	coauthorship	326,186	1,615,400	238	.3018	.2604	-.7736
(vii)	dblp-2011	coauthorship	986,324	6,707,236	979	.0842	.1351	-.2963
(viii)	hollywood-2009	costarring	1,139,905	113,891,327	11,468	.3446	.4689	-0.6737

**Table 5.** (i)–(iv) Web crawls: nodes are webpages, and an (undirected) edge means that there is a hyperlink from one of the two pages to another. (iii) and (iv) Breadth-first crawls: around one page. (v) E-mail exchange by Enron employees (mostly by the senior management): nodes are employees, and an edge means that an e-mail message was sent from one of the two employees to another. (vi), (vii) Scientific collaboration networks extracted from the DBLP bibliography service: each vertex represents a scientist and an edge means a coauthorship of at least one article. (viii) Vertices are actors: two actors are connected by an edge if they appeared in the same movie.

### 5.3. Web Samples and Social Networks

For completeness, we present the numerical results for web samples and social networks from [van der Hofstad and Litvak 13], see in Table 5. We used the compressed graph data from the Laboratory of Web Algorithms (LAW) at the Università degli studi di Milano [Boldi and Vigna 04, Boldi et al. 11] with the `bvgraph` MATLAB package [Gleich et al. 10]. The *stanford-cs* database [Constantine and Gleich 07] is a 2001 crawl that includes all pages in the `cs.stanford.edu` domain. In datasets (iv), (vii), and (viii) we evaluate  $\rho(G_n)$ ,  $\rho^{\text{rank}}(G_n)$ , and  $\rho^-(G_n)$  (see (3.5)) over 1000 random edges, and present the average over 10 such evaluations (in 10 samples of 1000 edges, the observed dispersion of the results was small).

We note that  $\rho^{\text{rank}}(G_n)$  here is an approximation of (3.2) computed as described in [van der Hofstad and Litvak 13]: we define the random variables  $X$  and  $Y$  as the degrees on two ends of a random *undirected* edge in a graph (that is, here  $(u, v)$  and  $(v, u)$  represent the same edge); for each edge, when the observed degrees are  $a$  and  $b$ , we assign  $[X = a, Y = b]$  or  $[X = b, Y = a]$  with probability  $1/2$ ; the ties are resolved randomly as in (3.2). The experiments on random graphs show that the values obtained by this algorithm are very close to those computed by (3.2).

The most remarkable result here is obtained on the two .uk crawls (iii) and (iv). Here  $\rho(G_n)$  is significantly smaller in magnitude on a larger crawl. Intuitively, mixing patterns should not depend on the crawl size. This is indeed confirmed by the value of Spearman's rho, which consistently shows strong negative correlations in both crawls. We could not observe a similar phenomenon so sharply in (vi) and (vii), probably because a larger coauthorship network incorporates articles from different areas of science, and the culture of scientific collaborations can vary greatly from one research field to another.

We also notice that, as predicted by our results, the size in magnitude values of  $\rho^-(G_n)$  result in profound difference in magnitude between  $\rho(G_n)$  and  $\rho^{\text{rank}}(G_n)$ . This is clearly seen in the data sets (ii), (iv), and (v). Again, (ii) and (iv) are the largest among the analyzed web crawls.

The observed behavior of Pearson's coefficient is explained by the results proved in this study in that  $\rho(G_n)$  is strongly influenced by the large dispersion in the degree values, and particularly by the presence of hubs. The latter increases with graph size because of the scale-free phenomenon. As a result,  $\rho(G_n)$  becomes smaller in magnitude when  $n$  increases, which makes it impossible to compare graphs of different sizes. In contrast, the *ranks* of the degrees are drawn from a uniform distribution on  $[0, 1]$ , scaled by the factor  $|E'|$ . Clearly, when a correlation coefficient is computed, the scaling factor cancels, and therefore, Spearman's rho provides consistent results in the graphs of different sizes.

## 6. Discussion

In this study, we have investigated dependency measures for power-law random variables. We have argued that Pearson's correlation coefficient, despite its appealing feature that it is always in  $[-1, 1]$ , is inappropriate to describe dependencies between heavy-tailed random variables. Indeed, the two main problems with the sample correlation coefficient are that (a) it can converge to a proper random variable when the sample size tends to infinity, indicating that it fluctuates tremendously as the sample size increases, and (b) it is always asymptotically nonnegative when dealing with nonnegative random variables (even when these are obviously negatively dependent). In the context of random graphs, the first deficiency means that Pearson's coefficient can have a nonvanishing variance even when the size of the graph is huge, the second deficiency mistakenly suggests that there do not exist asymptotically disassortative scale-free graphs. We give proofs for the facts stated above, and illustrate the results using simulations.

Rank correlations are a special case of the broader concept of copulas that are widely used in multivariate analysis, in particular in applications in mathematical finance and risk management. There is a heated discussion in this area about the adequacy and informativeness of such measures, see e.g. [Mikosch 06] and consequent reactions. There are several points of criticism. In particular, Spearman's rho uses rank transformation, which changes the observed values of the degrees. Then, first of all, what exactly does Spearman's rho tell us about the dependence between the original values? Second of all, no substantial justification exists for the rank transformation, besides its mathematical convenience. We thus do not claim that Spearman's rho is *the* solution to the problem. Nevertheless, compared to the Pearson's coefficient, Spearman's rho has a significant advantage, that it is free from the undesirable size-dependency, and converges to a meaningful value in the infinite volume limit.

We note that Spearman's rho has computational complexity  $O(n \log(n))$  because the values of the random variables must be ranked first. Pearson's correlation coefficient is easier to evaluate because it uses the values of the degrees directly, and has computational complexity  $O(n)$ . Efficient methods for computing Spearman's rho in large graphs is an interesting topic for future research.

Raising the discussion to a higher level, random variables  $X$  and  $Y$  are positively dependent when a large realization of  $X$  typically implies a large realization of  $Y$ . A strong form of this notion is when  $\mathbb{P}(X > x, Y > y) \geq \mathbb{P}(X > x)\mathbb{P}(Y > y)$  for every  $x, y \in \mathbb{R}$ , but for many purposes this notion is too restrictive. The covariance for nonnegative random variables is obtained by integrating the above inequality over  $x, y \geq 0$ , so that it is true for "typical" values

of  $x, y$ . In many cases, however, we are particularly interested in certain values of  $x, y$ . Another class of methods for measuring rank correlations is based on the *angular measure*, a notion originating in the theory of multivariate extremes, for which the above inequality is investigated for *large*  $x$  and  $y$ , so that it describes the *tail dependence* for a random vector  $(X, Y)$ , that is, the dependence between extremely large values of  $X$  and  $Y$ , see e.g. [Resnick 07]. Such tail dependence is characterized by probability-like measure, or, the angular measure, on  $[0, 1]$ . Informally, a concentration of the angular measure around the points 0 and 1 indicates independence of large values, although concentration around some other number  $a \in (0, 1)$  suggests that a certain fraction of large values of  $Y$  comes together with large values of  $X$ . In [Volkovich et al. 08, 09] a first attempt was made to compute the angular measure between in-degree of a node and its importance measured by the Google PageRank algorithm. Strikingly, completely different dependence structures were discovered in Wikipedia (independence), preferential attachment networks (complete dependence) and the Web (intermediate case).

**Acknowledgments.** We thank Yana Volkovich for the code generating a preferential attachment graph and Marie Albenque for a counterexample that shows that negative dependence of  $(X, Y)$  does not follow from negative dependence of  $(X_1, Y_1)$ . We further thank Juli Komjáthy for her careful reading of the paper, which has tremendously improved the presentation and has corrected several typos and errors. This article is also the result of joint research in the 3TU Centre of Competence NIRICT (Netherlands Institute for Research on ICT) within the Federation of Three Universities of Technology in The Netherlands.

**Funding.** The work of RvdH was supported in part by the Netherlands Organisation for Scientific Research (NWO). The work of NL is partially supported by the EU-FET Open grant NADINE (288956).

## References

- [Albert and Barabási 02] R. Albert and A. L. Barabási. “Statistical Mechanics of Complex Networks.” *Reviews of Modern Physics*, 74:1 (2002), 47.
- [Azuma 67] K. Azuma. “Weighted Sums of Certain Dependent Random Variables.” *Tohoku Mathematical Journal*, 19:3 (1967), 357–367.
- [Bender and Canfield 78] E.A. Bender and E.R. Canfield. “The Asymptotic Number of Labeled Graphs with Given Degree Sequences.” *Journal of Combinatorial Theory, Series A*, 24:3 (1978), 296–307.
- [Bingham et al. 89] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation of Encyclopedia of Mathematics and its Applications*, 27. Cambridge, UK: Cambridge University Press, 1989.
- [Boldi et al. 11] P. Boldi, M. Rosa, M. Santini, and S. Vigna. “Layered Label Propagation: A Multiresolution Coordinate-Free Ordering for Compressing Social

- Networks.” In *Proceedings of the 20th International World Wide Web Conference (WWW 2011)*. ACM Press, 2011.
- [Boldi and Vigna 04] P. Boldi and S. Vigna. “The WebGraph Framework I: Compression Techniques.” In *Proceedings of the 13th International World Wide Web Conference (WWW 2004)* pp. 595–601, Manhattan: ACM Press, 2004.
- [Bollobás 80] B. Bollobás. “A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs.” *European J. Combin.* 1:4 (1980), 311–316.
- [Bollobás 01] B. Bollobás. *Random Graphs* vol. 73. Cambridge, Cambridge, Cambridge University Press, 2001.
- [Bollobás et al. 01] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The Degree Sequence of a Scale-Free Random Graph Process. *Random Structures and Algorithms* 18 (2001), 279–290.
- [Borkowf 02] C. B. Borkowf. “Computing the Nonnull Asymptotic Variance and the Asymptotic Relative Efficiency of Spearman’s Rank Correlation.” *Computational Statistics & Data Analysis* 39:3 (2002), 271–286.
- [Braha and Bar-Yam 07] D. Braha and Y. Bar-Yam. “The Statistical Mechanics of Complex Product Development: Empirical and Analytical Results.” *Management Science* 53:7 (2007), 1127–1145.
- [Broder et al. 00] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Statac, A. Tomkins, and J. Wiener. “Graph Structure in the Web.” *Computer Networks* 33, (2000), 309–320.
- [Constantine and Gleich 07] P. G. Constantine and D. F. Gleich. “Using Polynomial Chaos to Compute the Influence of Multiple Random Surfers in the PageRank Model.” In edited by A. Bonato and F. C. Graham, pp. 82–95. *Proceedings of the 5th Workshop on Algorithms and Models for the Web Graph (WAW2007)*, Lecture Notes in Computer Science 4863, Berlin: Springer, 2007.
- [Dorogovtsev et al. 10] S. N. Dorogovtsev, A. L. Ferreira, A. V. Goltsev, and J. F. F. Mendes. “Zero Pearson Coefficient for Strongly Correlated Growing Trees.” *Physical Review E* 81:3 (2010), 031135.
- [Doyle et al. 05] J. C. Doyle, D. L. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger. “The Robust Yet Fragile Nature of the Internet.” *PNAS* 102:41 (2005), 14497–14502.
- [Durrett 07] R. Durrett. *Random Graph Dynamics*. Cambridge, UK: Cambridge University Press, 2007.
- [Eguiluz and Klemm 02] V. M. Eguiluz and K. Klemm. “Epidemic Threshold in Structured Scale-Free Networks.” *Physical Review Letters* 89:10 (2002), 108701.
- [Eubank et al. 04] S. Eubank, H. Guclu, V. S. Anil Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. “Modelling Disease Outbreaks in Realistic Urban Social Networks.” *Nature*, 429:6988 (2004), 180–184.
- [Fortunato et al. 07] S. Fortunato, M. Boguñá, A. Flammini, and F. Menczer. “On Local Estimations of PageRank: A Mean Field Approach.” *Internet Mathematics* 4: 2–3 (2007), 245–266.

- [Gleich et al. 10] D. F. Gleich, A. P. Gray, C. Greif, and T. Lau. “An Inner-Outer Iteration for Computing PageRank.” *SIAM Journal on Scientific Computing* 32:1 (2010), 349.
- [Gnedenko and Kolmogorov 68] B. V. Gnedenko and A. N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables*. Translated from the Russian, annotated, and revised by K. L. Chung. With appendices by J. L. Doob and P. L. Hsu. Revised edition. Addison-Wesley, Reading, MA; London-Don Mills., Ont: Addison-Wesley, 1968.
- [Hoeffding 63] W. Hoeffding. “Probability Inequalities for Sums of Bounded Random Variables.” *Journal of the American Statistical Association* 58:301 (1963), 13–30.
- [van der Hofstad 13] R. van der Hofstad. “Random Graphs and Complex Networks 2013. Available at <http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf>.
- [van der Hofstad and Litvak 13] R. van der Hofstad and N. Litvak. “Uncovering Disassortativity in Large Scale-Free Networks.” *Physical Review E* 87:2 (2013), 022801.
- [Janson 09] S. Janson. “The Probability that a Random Multigraph is Simple.” *Combinatorics, Probability and Computing*, 18:1–2 (2009), 205–225.
- [Kendall 75] M. Kendall. *Rank Correlation Methods*. London, UK: Charles Griffin & Company, 1975.
- [Kendall 38] M. G. Kendall. “A New Measure of Rank Correlation”. *Biometrika* 30:1-2 (1938), 81–93.
- [Kumar and Vassilvitskii 10] R. Kumar and S. Vassilvitskii. “Generalized Distances Between Rankings.” In *Proceedings of the 19th International Conference on World Wide Web*, pp. 571–580. New York, NY: ACM, 2010.
- [Li et al. 05] L. Li, D.L. Alderson, J.C. Doyle, and W. Willinger. “Towards a Theory of Scale-Free Graphs: Definition, Properties, and implications.” *Internet Mathematics* 2:4 (2005), 431–523.
- [Mahadevan et al. 06] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat. “Systematic Topology Analysis and Generation Using Degree Correlations.” *ACM SIGCOMM Computer Communication Review*, 36:4 (2006), 135–146.
- [Mesfioui and Tajar 05] M. Mesfioui and A. Tajar. “On the Properties of Some Non-parametric Concordance Measures in the Discrete Case.” *Nonparametric Statistics* 17:5 (2005), 541–554.
- [Mikosch 06] T. Mikosch. “Copulas: Tales and Facts.” *Extremes* 9:1 (2006), 3–20.
- [Mitzenmacher 04] M. Mitzenmacher. “A Brief History of Generative Models for Power Law and Lognormal distributions.” *Internet Mathematics* 1:2 (2004), 226–251.
- [Molloy and Reed 95] M. Molloy and B. Reed. “A Critical Point for Random Graphs with a Given Degree Sequence.” *Random Structures & Algorithms* 6:2–3 (1995), 161–180.
- [Molloy and Reed 98] M. Molloy and B. Reed. “The Size of the Giant Component of a Random Graph with a Given Degree Sequence.” *Combinatorics Probability and Computing* 7:3 (1998), 295–305.

- [Nevslehová 07] J. Nevslehová. “On Rank Correlation Measures for Non-Continuous Random Variables.” *Journal of Multivariate Analysis* 98:3 (2007), 544–567.
- [Newman 02] M. E. J. Newman. “Assortative Mixing in Networks.” *Physical Review Letters* 89:20 (2002), 208701.
- [Newman 03a] M. E. J. Newman. “Mixing Patterns in Networks.” *Physical Review E* 67:2 (2003), 026126.
- [Newman 03b] M. E. J. Newman. “The Structure and Function of Complex Networks.” *SIAM Review* 45:2 (2003), 167–256.
- [Newman 10] M. Newman. *Networks: An Introduction*. New York, NY: Oxford University Press, 2010.
- [Newman et al. 01] M. E. J. Newman, S.H. Strogatz, and D.J. Watts. “Random Graphs with Arbitrary Degree Distributions and Their Applications.” *Physical Review E* 64:2 (2001), 026118.
- [Raschke et al. 10] M. Raschke, M. Schläpfer, and R. Nibali. “Measuring Degree-Degree Association in Networks.” *Physical Review E* 82:3 (2010), 037102.
- [Resnick 07] S. I. Resnick. *Heavy-Tail Phenomena*. Berlin, Heidelberg: Springer, 2007.
- [Spearman 04] C. Spearman. “The Proof and Measurement of Association Between Two Things.” *The American Journal of Psychology* 15:1 (1904), 72–101.
- [Volkovich et al. 08] Y. Volkovich, N. Litvak, and B. Zwart. “Measuring Extremal Dependencies in Web Graphs.” In *WWW’ 08: Proceedings of the 17th International conference on World Wide Web*, pp. 1113–1114. New York, NY: ACM Press, 2008.
- [Volkovich et al. 09] Y. Volkovich, N. Litvak, and B. Zwart. “Extremal Dependencies and Rank Correlations in Power Law Networks.” In *Complex Sciences*, vol. 5, edited by J. Zhou, O. Akan, and P. Bellavista et al. pp. 1642–1653. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Berlin, Heidelberg: Springer, 2009.

---

Remco van der Hofstad, Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands (rhofstad@win.tue.nl)

Nelly Litvak, Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands (n.litvak@ewi.utwente.nl)