

Degree-degree correlations in random graphs with heavy-tailed degrees *

Nelly Litvak[†] and Remco van der Hofstad[‡]

University of Twente and Eindhoven University of Technology

April 2, 2012

Abstract

We investigate degree-degree correlations for scale-free graph sequences. The main conclusion of this paper is that the assortativity coefficient is *not* the appropriate way to describe degree-dependences in scale-free random graphs. Indeed, we study the infinite volume limit of the assortativity coefficient, and show that this limit is always non-negative when the degrees have finite first but infinite third moment, i.e., when the degree exponent $\gamma + 1$ of the density satisfies $\gamma \in (1, 3)$. More generally, our results show that the correlation coefficient is inappropriate to describe dependencies between random variables having infinite variance.

We start with a simple model of the sample correlation of random variables X and Y , which are linear combinations with non-negative coefficients of the *same* infinite variance random variables. In this case, the correlation coefficient of X and Y is not defined, and the sample covariance converges to a proper random variable with support that is a subinterval of $(-1, 1)$. Further, for any joint distribution (X, Y) with equal marginals being non-negative power-law distributions with infinite variance (as in the case of degree-degree correlations), we show that the limit is non-negative. We next adapt these results to the assortativity in networks as described by the degree-degree correlation coefficient, and show that it is non-negative in the large graph limit when the degree distribution has an infinite third moment. We illustrate these results with several examples where the assortativity behaves in a non-sensible way.

We further discuss alternatives for describing assortativity in networks based on *rank correlations* that are appropriate for infinite variance variables. We support these mathematical results by simulations.

Keywords. Dependencies of heavy-tailed random variables, Power-laws, Scale-free graphs, Assortativity, Degree-degree correlations, Multivariate extremes

1 Introduction

Large self-organizing networks, such as the Internet, the World Wide Web, social and biological networks, often exhibit power-law degrees. 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 . Power-law distributions are *heavy tailed* since the tail probability decreases much more slowly than negative exponential, and thus one observes extremely large values of X much more frequently than in the case of light tails. In the network context, such networks are called *scale free*, and the vertices having huge degrees are called hubs. Statistical analysis of complex networks characterized by power-law degrees has received massive attention in recent literature, see

*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.

[†]This research is done in the framework of the EC FET NADINE project.

[‡]This work was supported in part by the Netherlands Organisation for Scientific Research (NWO)

e.g. [10, 17, 22] for excellent surveys. Nevertheless, there still are many fundamental open problems. One of them is how to measure *dependencies* between network parameters.

An important property of networks is the dependence between the degrees of direct neighbours. Often, this dependence is characterized by the *assortativity coefficient* of the network, introduced by Newman in [20]. The assortativity coefficient is the correlation coefficient between the vector of degrees on each side of an edge, as a function of all the edges. See [20, Table I] for a list of assortativity coefficients for various real-world networks. The empirical data suggest that social networks tend to be assortative (i.e., the assortativity coefficient is positive), while technological and biological networks tend to be disassortative. In [20, Table I], it is striking that, typically, larger disassortative networks have an assortativity coefficient that is closer to 0 and therefore appear to have approximate *uncorrelated* degrees across edges. Similar conclusions can be drawn from [21], see in particular [21, Table II]. In this paper, we explain this effect mathematically, and conclude that the assortativity coefficient is *not* the way to describe dependencies between degrees on edges in the case of scale-free networks.

Instead, we propose a solution based on the *ranks* of degrees to deal with degree-degree dependencies in networks. This rank correlation approach is in fact standard and even classical in the area of 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, while the assortativity coefficient, despite the fact that it is always in $[-1, 1]$, suffers from a strong dependence on the extreme values of the degrees. This was already noted in the 1936 paper by H. Hotelling and M. R. Pabst [11]: ‘*Certainly where there is complete absence of knowledge of the form of the bivariate distribution, and especially if it is believed not to be normal, the rank correlation coefficient is to be strongly recommended as a means of testing the existence of relationship.*’ Among recent applications of rank correlation measures, such as Spearman’s rho [26] and the closely related Kendall’s tau [13], is measuring 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 e.g. [14] and references therein.

We will show in numerical experiments 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 1.000 and 100.000 nodes, respectively), these estimators give consistent values, and the variance of the estimator decreases as the size of the graph grows. We also analytically and numerically show that the assortativity coefficient absolutely does not have this basic property when degree distributions are heavy-tailed!

The paper is organized as follows. We start with formal definitions of the sample correlation coefficient and the sample rank correlation in Section 2. In Section 3 we study a model with linear dependencies and demonstrate that, when sample size grows to infinity, the sample correlation coefficient (assortativity) does not converge to a constant but rather to a random variable involving stable distributions. We also verify numerically that the rank correlation provides a consistent statistical estimator for this model. Next, in Section 4 we prove that if random variables are heavy-tailed and non-negative, then the sample correlation coefficient never converges to a negative value. Thus, such sequence will never be classified as ‘disassortative’. We illustrate this result by an example of two nonnegative but negatively correlated random variables. This result is extended to random graphs in Section 5, where also numerical results are provided for assortativity coefficient and rank correlations in three different Configuration Models, and a Preferential Attachment graph.

2 Correlations between random variables

2.1 Sample correlation coefficient

The assortativity coefficient is in fact a statistical estimator of a *correlation coefficient* ρ for the degrees on the two ends of an arbitrary edge in a graph. In this section we formally define such estimator. The correlation coefficient ρ for two random variables X and Y with $\text{Var}(X), \text{Var}(Y) < \infty$ is defined by

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

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

$$\rho_n = \frac{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{S_n(X)S_n(Y)}, \quad (2.1)$$

where

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

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

$$S_n^2(X) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad S_n^2(Y) = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2 \quad (2.2)$$

denote the *sample variances*.

It is well known that, again under the assumption that $\text{Var}(X), \text{Var}(Y) < \infty$, the estimator ρ_n of ρ is *consistent*, i.e.,

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

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 always hold a.s., for any sample, and, therefore, one might be tempted to always use the sample correlation coefficient ρ_n in (2.1). In this paper we investigate what happens to ρ_n when $\text{Var}(X), \text{Var}(Y) = \infty$, and show that the use of ρ_n in random graphs is uninformative, and it leads to deceptive behavior in the context of a linear dependence, such as in (3.1) below.

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 the two components of the resulting vector (r_i^X, r_i^Y) are realisations of identical uniform distributions, implying many nice mathematical properties.

The statistical correlation coefficient for the rank is known as Spearman's rho [26]:

$$\rho_n^{\text{rank}} = \frac{\sum_{i=1}^n (r_i^X - (n+1)/2)(r_i^Y - (n+1)/2)}{\sqrt{\sum_{i=1}^n (r_i^X - (n+1)/2)^2 \sum_{i=1}^n (r_i^Y - (n+1)/2)^2}} = 1 - \frac{6 \sum_{i=1}^n l_i^2}{n^3 - n}, \quad (2.3)$$

where $l_i = r_i^X - r_i^Y$, see [11]. The mathematical properties of the Spearman's rho have been extensively investigated in the literature. In particular, if $((X_i, Y_i))_{i=1}^n$ consists of independent realizations of (X, Y) , and the joint distribution function of X and Y is differentiable, then ρ_n^{rank} is a consistent estimator, and its standard deviation is of the order $1/\sqrt{n}$ (see e.g. [8], where exact asymptotic expressions are derived).

3 Linear dependencies

It is well known that ρ in general measures *linear* dependence between two random variables. Thus, it is natural to check how this measure, and our proposed alternative measures, perform when the relation between X and Y are described through the following linear model:

$$X = \alpha_1 U_1 + \cdots + \alpha_m U_m, \quad Y = \beta_1 U_1 + \cdots + \beta_m U_m, \quad (3.1)$$

where U_j , $j = 1, \dots, m$, are independent identically distributed (i.i.d.) random variables with regularly varying tail, and tail exponent γ . By definition, the random variable U is *regularly varying* with index $\gamma > 0$, if

$$\mathbb{P}(U > x) = \mathbb{P}(V > x) = L(x)x^{-\gamma}, \quad (3.2)$$

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

Theorem 3.1 (Weak convergence of correlation coefficient). *Let $((X_i, Y_i))_{i=1}^n$ be i.i.d. copies of the random variables (X, Y) in (3.1), and where $(U_j)_{j=1}^m$ are i.i.d. random variables satisfying (3.2) with $\gamma \in (0, 2)$, so that $\text{Var}(U_j) = \infty$. Then,*

$$\rho_n \xrightarrow{d} \rho \equiv \frac{\sum_{j=1}^m \alpha_j \beta_j Z_j}{\sqrt{\sum_{j=1}^m \alpha_j^2 Z_j} \sqrt{\sum_{j=1}^m \beta_j^2 Z_j}}, \quad (3.3)$$

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, ρ has a density on $[-1, 1]$, which is strictly positive on $(-1, 1)$ if there exist k, l such that $\alpha_k \beta_k < 0 < \alpha_l \beta_l$, while the density is positive on $(a, 1)$ when $\alpha_k \beta_k \geq 0$ for every k , where

$$a = \min_{S \subset \{1, 2, \dots, m\}, |S| \geq 2} \frac{\sum_{j=1}^m \alpha_j \beta_j \mathbf{1}_{\{j \in S\}}}{\sqrt{\sum_{j=1}^m \alpha_j^2 \mathbf{1}_{\{j \in S\}}} \sqrt{\sum_{j=1}^m \beta_j^2 \mathbf{1}_{\{j \in S\}}} \in (0, 1). \quad (3.4)$$

In order to prove the theorem we need the following technical result:

Lemma 3.2 (Asymptotics of sums in stable domain). *Let $(U_{i,j})_{i \in [n], j \in [2]}$ be i.i.d. random variables satisfying (3.2) 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*

$$\frac{1}{a_n} \sum_{i=1}^n U_{i,1}^2 \xrightarrow{d} Z_1, \quad \frac{1}{a_n} \sum_{i=1}^n U_{i,1} U_{i,2} \xrightarrow{\mathbb{P}} 0, \quad (3.5)$$

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

Proof. Denote by F the distribution function of U . The proof of the first statement in (3.5) is classical when we note that the distribution function of U^2 equals $u \mapsto F(\sqrt{u})$, which, by (3.2), is in the domain of attraction of a stable $\gamma/2$ random variable. In particular, we can identify $a_n = [1 - F]^{-1}(1/n^2)$. To prove the second part of (3.5), we write

$$1 - F(x) = \mathbb{P}(U > x) \leq c' x^{-\gamma'}, \quad x \geq 0, \quad (3.6)$$

which is valid for any $\gamma' \in (1, \gamma)$ by (3.2) and Potter's theorem. We next study the distribution function of $U_1 U_2$ which we denote by H , where U_1 and U_2 are two independent copies of the random variable U . When F satisfies (3.6), then it is not hard to see that there exists a $C > 0$ such that

$$1 - H(u) \leq C(1 + \log u)u^{-\gamma'}. \quad (3.7)$$

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 (3.6), then U_1 and U_2 are stochastically upper bounded by U_1^* and U_2^* with distribution function F^* satisfying $1 - F^*(w) = c'w^{-\gamma'} \vee 1$, where $(x \vee y) = \max\{x, y\}$, and the claim in (3.7) follows from the above computation.

By the bound in (3.7), the random variables $U_{i,1}U_{i,2}$ are stochastically bounded from above by random variables P_i that are in the domain of attraction of a stable γ' 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 γ' . By choosing $\gamma' > \gamma/2$, we get $b_n/a_n \rightarrow 0$, so we obtain the second statement in (3.5) \square

Proof of Theorem 3.1. We start by noting that

$$\rho_n = \frac{\frac{1}{n-1} \sum_{i=1}^n (X_i Y_i - \bar{X}_n \bar{Y}_n)}{S_n(X) S_n(Y)}, \quad (3.8)$$

and

$$S_n^2(X) = \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - \bar{X}_n^2), \quad S_n^2(Y) = \frac{1}{n-1} \sum_{i=1}^n (Y_i^2 - \bar{Y}_n^2). \quad (3.9)$$

We continue to identify the asymptotic behavior of

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

The distribution of $((X_i, Y_i))_{i=1}^n$ is described in terms of an array $(U_{i,j})_{i \in [n], j \in [m]}$, which are i.i.d. copies of a random variable U . In terms of these random variables, we can identify

$$\sum_{i=1}^n X_i Y_i = \sum_{j=1}^m \alpha_j \beta_j \left(\sum_{i=1}^n U_{i,j}^2 \right) + \sum_{j_1 \neq j_2=1}^m \alpha_{j_1} \beta_{j_2} \left(\sum_{i=1}^n U_{i,j_1} U_{i,j_2} \right). \quad (3.10)$$

The sums $\sum_{i=1}^n U_{i,j}^2$ are i.i.d., and by Lemma 3.2, $\sum_{i=1}^n U_{i,j_1} U_{i,j_2}$ is of a smaller order. Hence, from (3.10) we obtain that

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

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

$$\frac{1}{a_n} \sum_{i=1}^n X_i^2 \xrightarrow{d} \sum_{j=1}^m \alpha_j^2 Z_j, \quad \frac{1}{a_n} \sum_{i=1}^n Y_i^2 \xrightarrow{d} \sum_{j=1}^m \beta_j^2 Z_j, \quad (3.11)$$

and these convergence hold *simultaneously*. As a result, (3.3) follows. It remains to establish properties of the limiting random variable ρ in (3.3).

The density of Z_i is strictly positive on $(0, \infty)$, so that the density of ρ is strictly positive on $(-1, 1)$ when the sign of $\alpha_i \beta_i$ is both positive as well as negative. When $\alpha_i \beta_i \geq 0$ for every i , on the other hand, the density of ρ is strictly positive on the support of ρ , which is $(a, 1)$, where

$$a = \inf_{z_1, \dots, z_m} \frac{\sum_{j=1}^m \alpha_j \beta_j z_j}{\sqrt{\sum_{j=1}^m \alpha_j^2 z_j} \sqrt{\sum_{j=1}^m \beta_j^2 z_j}} \in (0, 1). \quad (3.12)$$

Denote the function that is minimized by $a(z_1, \dots, z_m)$. Note that rescaling $z_j = cz_j$, $j = 1, \dots, m$, does not change the value of $a(z_1, \dots, z_m)$. In particular, we can choose $c = (\max\{z_1, z_2, \dots, z_m\})^{-1}$. Thus, without loss of generality, we can assume that $z_j \in (0, 1]$, $j = 1, 2, \dots, m$ and $z_k = 1$ for at least one $k = 1, 2, \dots, m$. In that case, a is a continuous function of $z_j \in [0, 1]$, $j \neq k$. Taking a derivative of a with respect to z_j we obtain that the sign of the derivative is defined by the sign of the expression

$$-a(z_1, \dots, z_m)(\alpha_j^2 + \beta_j^2) + 2\alpha_j \beta_j. \quad (3.13)$$

Since (3.13) is decreasing in a , the derivative of $a(z_1, \dots, z_m)$ w.r.t. z_j cannot equal zero. (Indeed, if (3.13) is zero in some point z_j^* , then (3.13) is *positive* on $(z_j^*, z_j^* + \varepsilon)$ for some small ε only if a is *decreasing* on $(z_j^*, z_j^* + \varepsilon)$, thus, we obtain a contradiction.) We conclude that a achieves its minimum when all z_j 's equal either zero or one, and at least one of the values must equal one. Finally, if only one value z_j is equal to one and the rest are equal to zero, then we obtain $a = 1$, which is a maximal possible value of a . Thus, at least two values of z_j must equal one. \square

To illustrate the result of Theorem 3.1, consider the example with U_j 's from a Pareto distribution satisfying $\mathbb{P}(U > x) = 1/x^{1.1}$, $x \geq 1$, so $L(x) = 1$ and $\gamma = 1.1$ in (3.2). The exponent $\gamma = 1.1$ is as observed for the World Wide Web [9]. In (3.1), we choose $m = 3$ and α_i, β_i , $i = 1, 2, 3$, as specified in Table 1. We generate N data samples $((X_i, Y_i))_{i=1}^n$ and compute ρ_n and ρ_n^{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 realisations for ρ_n and ρ_n^{rank} , respectively, where the sub-index $j = 1, \dots, N$ denotes the j th realization of $((X_i, Y_i))_{i=1}^n$. We then compute

$$\mathbb{E}_N(\rho_n) = \frac{1}{N} \sum_{j=1}^N \rho_{n,j}, \quad \mathbb{E}_N(\rho_n^{\text{rank}}) = \frac{1}{N} \sum_{j=1}^N \rho_{n,j}^{\text{rank}}, \quad (3.14)$$

$$\sigma_N(\rho_n) = \sqrt{\frac{1}{N-1} \sum_{j=1}^N (\rho_{n,j} - \mathbb{E}_N(\rho_n))^2}, \quad \sigma_N(\rho_n^{\text{rank}}) = \sqrt{\frac{1}{N-1} \sum_{j=1}^N (\rho_{n,j}^{\text{rank}} - \mathbb{E}_N(\rho_n^{\text{rank}}))^2}. \quad (3.15)$$

The results are presented in Table 1. We clearly see that ρ_n has a significant standard deviation, of which estimators are similar for different values of n . This means that in the limit as $n \rightarrow \infty$, ρ_n is a random variable with a significant spread in its values, as stated in Theorem 3.1. Thus, by evaluating ρ_n for one sample $((X_i, Y_i))_{i=1}^n$ we will obtain a random number, even when n is huge. The convergence to a non-trivial 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.

On the other hand, from Table 1 we clearly see that the behaviour of rank correlations is exactly as we can expect from a good statistical estimator. The obtained average values are consistent while the standard deviation of ρ_n^{rank} decreases approximately as $1/\sqrt{n}$ as n grows large. Therefore, ρ_n^{rank} converges to a deterministic number.

Model parameters	N	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 ρ_n and ρ_n^{rank} in N samples with linear dependence (3.1).

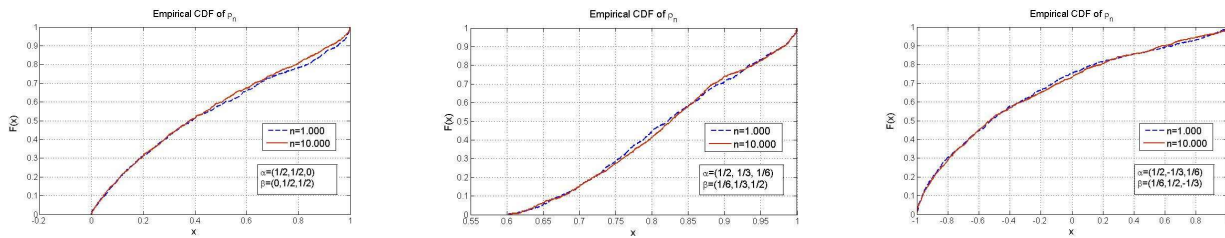


Figure 1: The empirical distribution function $F_N(x) = \mathbb{P}(\rho_n \leq x)$ for the $N = 1,000$ observed values of ρ_n ($n = 1,000$, $n = 10,000$), in the case of linear dependence (3.1).

4 Sample correlation coefficient for non-negative variables

In this section, we investigate correlations between non-negative heavy-tailed random variables. Our main result in this section shows that the correlation coefficient is asymptotically non-negative:

Theorem 4.1 (Asymptotic non-negativity of correlation coefficient for positive r.v.'s). *Let $((X_i, Y_i))_{i=1}^n$ be i.i.d. copies of non-negative random variables (X, Y) , where X and Y satisfy*

$$\mathbb{P}(X > x) = L_X(x)x^{-\gamma_X}, \quad \mathbb{P}(Y > y) = L_Y(y)y^{-\gamma_Y}, \quad x, y \geq 0, \quad (4.1)$$

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

We illustrate Theorem 4.1 with a useful example. Let $(U_i)_{i=1}^n$ be a sequence of i.i.d. random variables satisfying (3.2) for some $\gamma \in (0, 2)$, and where $U \geq 0$ a.s. Let $(X, Y) = (0, 2U)$ with probability $1/2$ and $(X, Y) = (2U, 0)$ with probability $1/2$. Then, $XY = 0$ a.s., while $\mathbb{E}[X] = \mathbb{E}[Y] = \mathbb{E}[U]$ and $\text{Var}(X) = \text{Var}(Y) = 2\mathbb{E}[U^2] - \mathbb{E}[U]^2 = 2\text{Var}(U) + \mathbb{E}[U]^2$. Therefore, if $\text{Var}(U) < \infty$,

$$\rho_n \xrightarrow{\mathbb{P}} \rho = -\frac{\mathbb{E}[U]^2}{2\text{Var}(U) + \mathbb{E}[U]^2} \in (-1, 0). \quad (4.2)$$

The asymptotics in (4.2) is quite reasonable, since the random variables (X, Y) are highly negatively dependent: When $X > 0$, Y must be equal to 0, and vice versa. Instead, when $(U_i)_{i=1}^n$ is a sequence of i.i.d. random variables satisfying (3.2) for some $\gamma \in (0, 2)$, and where $U \geq 0$ a.s., then $\rho_n \xrightarrow{\mathbb{P}} 0$, which is not what we would like.

Table 2 shows the empirical mean and standard deviation of the estimators ρ_n and ρ_n^{rank} . Here $\mathbb{P}(U > x) = x^{-1.1}$, $x \geq 1$, as in Table 1. As predicted by Theorem 4.1, the sample correlation coefficient (assortativity) converges to zero as n grows large, while ρ_n^{rank} consistently shows a clear

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 ρ_n and ρ_n^{rank} in N simulations of $((X_i, Y_i))_{i=1}^n$, where $X = 2UI$, $Y = 2U(1 - I)$, I is a Bernoulli(1/2) random variable, $\mathbb{P}(U > x) = x^{-1.1}$, $x \geq 1$.

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 power-law data.

We next prove Theorem 4.1:

Proof of Theorem 4.1. 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 show 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, and from (4.1), (2.2) and Lemma 3.5 it follows that $S_n(X) = n^{2/\gamma_X + o(1)-1} \rightarrow \infty$ as $n \rightarrow \infty$. When $\gamma \in (0, 1]$, instead, $\bar{X}_n = n^{1/\gamma_X - 1 + o(1)}$, so that $\bar{X}_n/S_n(X) = n^{-1/\gamma_X + o(1)} \xrightarrow{\mathbb{P}} 0$. This proves the claim. \square

5 Applications to random graph models

The correlation coefficient is particularly important in the setting of degree-degree correlations in real-world networks. Let $G = (V, E)$ be a graph with vertex set V and edge set E . The assortativity coefficient of G is equal to (see, e.g., [20, (4)])

$$\rho(G) = \frac{\frac{1}{|E|} \sum_{ij \in E} D_i D_j - \left(\frac{1}{|E|} \sum_{ij \in E} \frac{1}{2} (D_i + D_j) \right)^2}{\frac{1}{|E|} \sum_{ij \in E} \frac{1}{2} (D_i^2 + D_j^2) - \left(\frac{1}{|E|} \sum_{ij \in E} \frac{1}{2} (D_i + D_j) \right)^2},$$

where the sum is over directed edges of G , and D_i is the degree of vertex i , i.e., ij and ji are two distinct edges. The assortativity coefficient is equal to the correlation coefficient of the sequence of random variables $((D_i, D_j))_{ij \in E}$. Thus, the assortativity coefficient is the correlation coefficient between two sequences of non-negative random variables, as studied in Theorem 4.1. We refer to [23] for an extensive introduction to networks, their empirical properties and models for them.

This section is organized as follows. In Section 5.1 we show that all limit points of the assortativity coefficients for sequences of growing scale-free random graphs with power-law exponent $\gamma < 3$ are non-negative, a result that is similar in spirit to Theorem 4.1. We highlight this statement by presenting theoretical and numerical results for several random graph examples where the assortativity coefficient yields unexpected and unwanted results. In Section 5.2 we present an example of a sequence of random graphs where the assortativity coefficient converges to a proper random variable, as observed in the i.i.d. setting in Theorem 3.1.

5.1 No disassortative scale-free random graph sequences

We compute that

$$\frac{1}{|E|} \sum_{ij \in E} \frac{1}{2} (D_i + D_j) = \frac{1}{|E|} \sum_{i \in V} D_i^2, \quad \frac{1}{|E|} \sum_{ij \in E} \frac{1}{2} (D_i^2 + D_j^2) = \frac{1}{|E|} \sum_{i \in V} D_i^3.$$

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

$$\rho(G) = \frac{\sum_{ij \in E} D_i D_j - \frac{1}{|E|} \left(\sum_{i \in V} D_i^2 \right)^2}{\sum_{i \in V} D_i^3 - \frac{1}{|E|} \left(\sum_{i \in V} D_i^2 \right)^2}. \quad (5.1)$$

Consider a sequence of graphs $(G_n)_{n \geq 1}$, where n denotes the number of vertices $n = |V|$ 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 4.1 to this setting:

Theorem 5.1 (Asymptotic non-disassortativity of 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 $cn \leq |E| \leq Cn$, $cn^{1/\gamma} \leq \max_{i \in [n]} D_i \leq Cn^{1/\gamma}$ and $cn^{(2/\gamma) \vee 1} \leq \sum_{i \in [n]} D_i^2 \leq Cn^{(2/\gamma) \vee 1}$. Then, any limit point of the assortativity coefficients $\rho(G_n)$ is non-negative.*

Proof. We note that $D_i \geq 0$ for every $i \in V$, so that, from (5.1)

$$\rho(G_n) \geq \underline{\rho}(G_n) \equiv - \frac{\frac{1}{|E|} \left(\sum_{i \in V} D_i^2 \right)^2}{\sum_{i \in V} D_i^3 - \frac{1}{|E|} \left(\sum_{i \in V} D_i^2 \right)^2}.$$

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

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

This proves the claim. \square

In the literature, many examples are reported of real-world networks where the degree distribution obeys a power law (see [1, 22] for surveys of real-world networks and their degree properties). In particular, for scale-free networks, the proportion p_k of vertices of degree k is close to $p_k \approx ck^{-\gamma-1}$, and most values of γ reported in the literature are in $(1, 3)$, see e.g., [1, Table I] or [22, Table I]. When this is the case, we can expect that

$$|E| = \sum_{i \in V} D_i \sim \mu n,$$

where $\mu = \mathbb{E}[D]$, while $\max_{i \in V} D_i \sim n^{1/\gamma}$, and

$$\frac{1}{n} \sum_{i \in V} D_i^p \sim \begin{cases} \mu_p & \text{when } \gamma > p, \\ cn^{p/\gamma-1} & \text{when } \gamma < p, \end{cases}$$

where $\mu_p = \mathbb{E}[D^p]$. In particular, the conditions of Theorem 5.1 hold and $\underline{\rho}(G_n) \rightarrow 0$ when $\gamma < 3$. Thus, the asymptotic degree-degree correlation of the graph sequence $(G_n)_{n \geq 1}$ is non-negative. As a result, there exist no disassortative scale-free graph sequences.

We next consider four random graph models to highlight our result. In the remainder of this section we first describe three models: Configuration model, Configuration model with intermediate vertices, and Preferential Attachment model. Then we present the numerical results for these models in Table 3. As we see from the results, in all these models assortativity converges to zero as n grows.

The configuration model. The *configuration model* was invented by Bollobás in [5], inspired by [4]. Its connectivity structure was first studied by Molloy and Reed [18, 19]. It was popularized by Newman, Sroogatz and Watts [24], 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 and d_i half-edges adjacent to vertex i . The graph is constructed by randomly pairing each half-edge to some other half-edge to form an edge. Number the half-edges from 1 to ℓ_n in some arbitrary order. Then, at each step, two half-edges that are not already paired are chosen uniformly at random among all the unpaired half-edges and are paired to form a single edge in the graph. These half-edges are removed from the list of unpaired half-edges. We continue with this procedure of choosing and pairing two unpaired half-edges until all the half-edges are paired. Although self-loops may occur, these become rare as $n \rightarrow \infty$ (see e.g. [6] or [12] for more precise results in this direction). We consider both the cases where the self-loops are removed and we collapse multiple edges to a single edge, as well as the setting where we keep the self-loops and multiple edges. 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 are directly connected is close to $d_i d_j / \ell_n$. Since this is of product form in i and j , 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.

Configuration model with intermediate vertices. We next adapt the configuration model slightly, by replacing every edge by two edges that meet at a middle vertex. Denote this graph by $\bar{G}_n = (\bar{V}_n, \bar{E}_n)$, while the configuration model is $G_n = (V_n, E_n)$. In this model, there are $n + \ell_n/2$ vertices and $2\ell_n$ edges (recall that ij and ji are two different edges). For $st \in \bar{E}_n$, the degree of either vertex s or vertex t equals 2, and the degree of the other vertex in the edge is equal to D_i , where i is the unique vertex in the original configuration model that corresponds to s or t . Therefore,

$$\frac{1}{|\bar{E}_n|} \sum_{st \in \bar{E}_n} \bar{D}_s \bar{D}_t = \frac{2}{\ell_n} \sum_{i \in V_n} D_i^2,$$

and for $p \geq 2$,

$$\frac{1}{|\bar{E}_n|} \sum_{s \in \bar{V}_n} \bar{D}_s^p = \frac{1}{2\ell_n} 2^p (\ell_n/2) + \frac{1}{2\ell_n} \sum_{i \in V_n} D_i^p = 2^{p-2} + \frac{1}{2\ell_n} \sum_{i \in V_n} D_i^p,$$

where $\mu_p = \mathbb{E}[D^p]$. As a result, when $\gamma > 3$,

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

The fact that the degree-degree correlation is negative is quite reasonable, since in this model, vertices of high degree are only connected to vertices of degree 2, so that there is negative dependence between the degrees at either end of an edge. When $\gamma < 3$, on the other hand, $\mu_3 = \mathbb{E}[D^3] = \infty$, and thus

$$\rho(\bar{G}_n) \xrightarrow{\mathbb{P}} 0,$$

which is inappropriate, as the negative dependence of the degrees persists.

Preferential Attachment model. We consider the basic version of the undirected Preferential Attachment model (PAM), where each new vertex adds only one edge to the network, connecting to the existing nodes with probability proportional to their degrees. In this case, it is well known

that $\gamma = 2$ (see e.g., [2] or [7]). We see that the assortativity converges to zero, as indicated in Theorem 5.1, while Spearman's rank correlation indicates that degrees are negatively dependent. This can be understood by noting that the majority of edges of vertices with high degrees, which are old vertices, come from vertices which are added late in the graph growth process and thus have small degree. On the other hand, by the growth mechanism of the PAM, vertices with low degree are more likely to be connected to vertices having high degree, which indeed suggests negative degree-degree dependencies.

Numerical results. To illustrate our results, we have generated the two configuration models and the Preferential Attachment model of different sizes. For fair comparison, we chose $\gamma = 3$ for the configuration model: $\mathbb{P}(D \geq x) = x^{-2}$, $x \geq 1$. Since in the configuration model self-loops and multiple edges are possible, we considered 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 using (2.3) as follows. We define the random variables X and Y as the degrees on two ends of a random *undirected* edge in a graph (that is, when rank correlations are computed, ij and ji 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 $1/2$. Furthermore, many values of X and Y will be the same because a degree d will appear at the end of an edge d times. We resolve the draws by adding independent random variables, uniformly distributed on $[0, 1]$, to each value of X and Y . The results are presented in Table 3.

Model	N	10^3	10^2		10
	n	10^2	10^3	10^4	10^5
Configuration model with self-loops and double edges	$\mathbb{E}_N(\rho(G_n))$	0.0021	-0.0013	0.0001	-0.0003
	$\sigma_N(\rho(G_n))$	0.0672	0.0212	0.0068	0.0024
	$\mathbb{E}_N(\rho_n^{\text{rank}})$	0.0012	-0.0010	-0.0002	-0.0002
	$\sigma_N(\rho_n^{\text{rank}}(G_n))$	0.0656	0.0202	0.0066	0.0014
Configuration model without self-loops and double edges	$\mathbb{E}_N(\rho(G_n))$	-0.0785	-0.0346	-0.0115	-0.0046
	$\sigma_N(\rho(G_n))$	0.0686	0.0274	0.0102	0.0039
	$\mathbb{E}_N(\rho_n^{\text{rank}}(G_n))$	-0.0615	-0.0151	-0.0040	-0.0002
	$\sigma_N(\rho_n^{\text{rank}}(G_n))$	0.0836	0.0337	0.0075	0.0024
Configuration model with intermediate vertices	$\mathbb{E}_N(\rho(G_n))$	-0.2589	-0.1243	-0.0587	-0.0303
	$\sigma_N(\rho(G_n))$	0.0872	0.0509	0.0255	0.0189
	$\mathbb{E}_N(\rho_n^{\text{rank}}(\tilde{G}_n))$	-0.7482	-0.7499	-0.7498	-0.7501
	$\sigma_N(\rho_n^{\text{rank}}(\tilde{G}_n))$	0.0121	0.0036	0.0011	0.0006
Preferential attachment	$\mathbb{E}_N(\rho(G_n))$	-0.2597	-0.1302	-0.0607	-0.0294
	$\sigma_N(\rho(G_n))$	0.0550	0.0261	0.0127	0.0088
	$\mathbb{E}_N(\rho_n^{\text{rank}}(G_n))$	-0.4167	-0.4151	-0.4166	-0.4158
	$\sigma_N(\rho_n^{\text{rank}}(G_n))$	0.0695	0.0202	0.0066	0.0022

Table 3: Estimated mean and standard deviation of $\rho(G)$ and $\rho^{\text{rank}}(G)$ in random graphs.

Within the same model, the graphs of different sizes are constructed by the same algorithm. Thus, their mixing patterns are exactly the same. As we predicted, the assortativity reduces in absolute value with the graph size, resulting in asymptotically neutral mixing for all models. On the contrary, the rank correlation coefficient consistently shows neutral mixing for the configuration model, moderately disassortative mixing for the Preferential Attachment graph, and strongly disassortative mixing for the configuration model with intermediate edges.

5.2 Random graphs with asymptotically random assortativity

Finally, we discuss the possibility that $\rho(G_n)$ in (5.1) converges to a random variable when the number of vertices tends to infinity. Under the assumptions of Theorem 5.1, we have that

$$\sum_{ij \in E} D_i D_j \leq \max_{i \in [n]} D_i \sum_{ij \in E_n} D_i = \max_{i \in [n]} D_i \left(\sum_{i \in V_n} D_i^2 \right) \leq C^2 n^{1/\gamma + (2/\gamma \vee 1)}, \quad (5.2)$$

$$\sum_{ij \in E} D_i D_j \geq \max_{i \in [n]} D_i \geq cn^{1/\gamma}, \quad (5.3)$$

$$\sum_{ij \in E} D_i D_j \geq \sum_{i \in V_n} D_i^2 \geq cn^{2/\gamma \vee 1}. \quad (5.4)$$

Further, from the proof of Theorem 5.1, we know that

$$\sum_{i \in V} D_i^3 \geq (\max_{i \in [n]} D_i)^3 \geq c^3 n^{3/\gamma}, \quad (5.5)$$

and

$$\frac{1}{|E|} \left(\sum_{i \in V} D_i^2 \right)^2 \leq (C^2/c) n^{(4/\gamma - 1) \vee 1}, \quad (5.6)$$

where we see that (5.6) is vanishing compared to (5.5). The convergence of (5.1) to a random variable can only take place if the crossproducts on the left-hand side of (5.2 – 5.4) are of the same order of magnitude as the left-hand side of (5.5). As we see from the above, this is possible for $\gamma \in (1, 3)$. However, the convergence will be slow because an easy calculation shows that the maximal difference in the order of magnitude between the right-hand sides of (5.2) and (5.6) is $n^{1/2}$.

Below we present an example where we prove that $\rho(G_n)$ indeed converges to a random variable, and illustrate numerically how the distribution of $\rho(G_n)$ changes as n grows large. However, due to the slow convergence, a substantially larger computational capacity is needed in order to (almost) achieve the limiting distribution.

A collection of complete bipartite graphs. Take $((X_i, Y_i))_{i=1}^n$ to be an i.i.d. sample of random variables as in (3.1), 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 collection of n bipartite graphs by G_n . The graph G_n has $|V| = \sum_{i=1}^n (X_i + Y_i)$ vertices and $|E| = 2 \sum_{i=1}^n X_i Y_i$ edges. Further,

$$\sum_{i \in V} D_i^p = \sum_{i=1}^n (X_i^p Y_i + Y_i^p X_i), \quad \sum_{ij \in E} D_i D_j = 2 \sum_{i=1}^n (X_i Y_i)^2.$$

Assume that the U_j satisfy (3.2) with $\gamma \in (3, 4)$, so that $\mathbb{E}[U^3] < \infty$, but $\mathbb{E}[U^4] = \infty$. As a result, $|E|/n \xrightarrow{\mathbb{P}} 2\mathbb{E}[XY] < \infty$ and $\frac{1}{n} \sum_{i \in V} D_i^2 \xrightarrow{\mathbb{P}} \mathbb{E}[XY(X+Y)] < \infty$. Further,

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

where Z_1 and Z_2 and two independent stable distributions with parameter $\gamma/4$. As a result,

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

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

n	10^2	10^3	10^4	10^5
$\mathbb{E}_N(\rho(G_n))$	0.6855	0.7293	0.7877	0.8224
$\sigma_N(\rho(G_n))$	0.1389	0.0681	0.0614	0.0629
$\mathbb{E}_N(\rho^{\text{rank}}(G_n))$	0.7556	0.8370	0.8577	0.8641
$\sigma_N(\rho^{\text{rank}}(G_n))$	0.0791	0.0379	0.0247	0.0128

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 $N = 100$.

In Table 4 we present numerical results for $\rho(G_n)$ and $\rho^{\text{rank}}(G_n)$. Here we choose $b = 1/2$, $a = 2$, and U has a generalized Pareto distribution $\mathbb{P}(U > x) = ((1.8 + x)/2.8)^{-2.8}$, $x > 1$.

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 probability mass is spread over the interval $(0.8, 1)$.

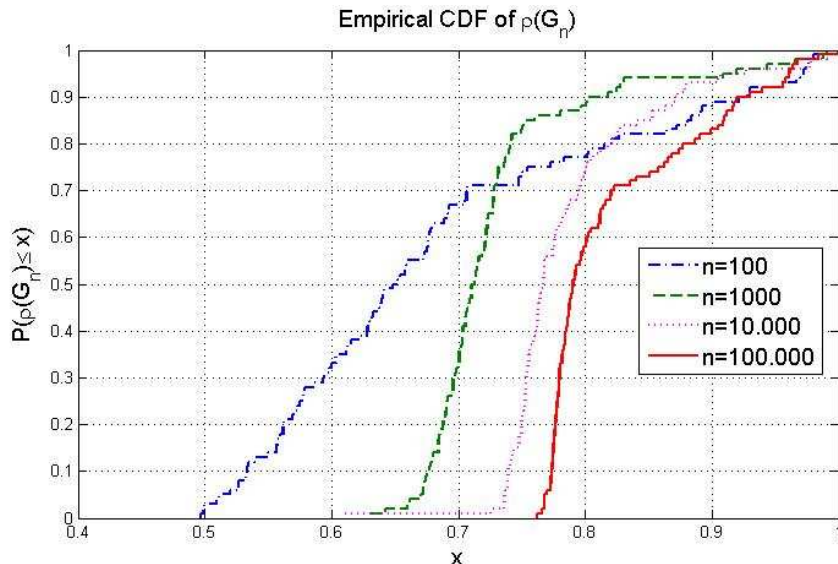


Figure 2: The empirical distribution function $\mathbb{P}(\rho(G_n) \leq x)$ for the $N = 100$ observed values of $\rho(G_n)$, where G_n is a collection of n complete bipartite graphs.

In the limit, $\rho(G_n)$ has a non-zero density on this interval. The difference between the crossproducts and the expectation squared in $\rho(G_n)$ is only of the order $n^{1/\gamma}$, which is $n^{1/2.8}$ in our example, thus, the convergence is too slow to observe it at $n = 100,000$.

6 Discussion

In this paper, we have investigated dependency measures for power-law random variables. We have argued that the correlation coefficient, despite its appealing feature that it is always in $[-1, 1]$, is inappropriate to describe dependencies between heavy-tailed random variables since it yields insensible results. 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 for different samples, and (b) that it is always asymptotically non-negative when dealing with non-negative random variables (even when these are obviously negatively dependent).

dent). In the context of random graphs, the first deficiency means that the assortativity can have a non-vanishing variance even when the size of the graph is huge, the second means that there do not exist asymptotically disassortative scale-free graphs. We give proofs for the facts stated above, and illustrate the results using simulations.

Further research is needed to study rank correlations on graphs. Although the numerical results suggest consistency of the Spearman’s rho estimator on random graphs, the values of the vectors $(X_i, Y_i)_{i \in E}$ are in general not independent. Indeed, if one edge emanates from a node with degree 250, then there are 499 other edges (ij and ji being different edges) for which either X_i or Y_i is equal to 250. Thus, the consistency of the estimator does not generally follow and needs a case by case proof.

Rank correlations are a special case of the broader concept of copulas that are widely used in multivariate analysis, in particular in application 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. [16] 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 for the problem. The main point of this paper is rather that the assortativity coefficient is *not* a solution at all, and that better solutions must be sought and can be found.

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 non-negative 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. [25]. Such tail dependence is characterized by an *angular measure* on $[0, 1]$. Informally, a concentration of the angular measure around the points 0 and 1 indicates independence of large values, while 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 [27, 28] 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).

Acknowledgment

We thank Yana Volkovich for the code generating a Preferential Attachment graph.

References

- [1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Modern Phys.*, **74**(1):47–97, (2002).
- [2] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, **286** (5439):509512, (1999).
- [3] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes: Theory and Applications*. Wiley, 2004.

- [4] E.A. Bender and E.R. Canfield. The asymptotic number of labelled graphs with a given degree sequences. *Journal of Combinatorial Theory (A)*, **24**:296–307, (1978).
- [5] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, **1**(4):311–316, (1980).
- [6] B. Bollobás. *Random graphs*, volume **73** of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, (2001).
- [7] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, **18**(3):279–290, 2001.
- [8] C. B. Borkowf. Computing the nonnull asymptotic variance and the asymptotic relative efficiency of Spearman’s rank correlation. *Computational Statistics & Data Analysis*, **39**(3): 271–286, 2002.
- [9] 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**(1-6): 309–320, 2000.
- [10] D. Chakrabarti and C. Faloutsos. Graph mining: Laws, generators, and algorithms. *ACM Comput. Surv.*, **38**(1):2, 2006.
- [11] H. Hotelling and M.R. Pabst. Rank correlation and tests of significance involving no assumption of normality. *The Annals of Mathematical Statistics*, **7**(1):29–43, 1936.
- [12] S. Janson. The probability that a random multigraph is simple. *Combinatorics, Probability and Computing*, **18**(1-2):205–225, 2009.
- [13] M. G. Kendall. A new measure of rank correlation. *Biometrika*, **30**(1/2): 81–93, 1938.
- [14] R. Kumar, R. and S. Vassilvitskii. Generalized distances between rankings. *Proceedings of the 19th international conference on World Wide Web*, pp. 571–580, ACM, 2010.
- [15] N. Litvak, W. R. W. Scheinhardt, Y. Volkovich, and B. Zwart. Characterization of Tail Dependence for In-Degree and PageRank. In K. Avrachenkov, D. Donato, and N. Litvak, editors, *Proceedings 6th International Workshop, WAW 2009, Barcelona, Spain*, volume 5427 of *Lecture Notes in Computer Science*, pages 90–103. Springer-Verlag Berlin, Heidelberg, 2009.
- [16] T. Mikosch. Copulas: Tales and facts., *Extremes*, **9**(1):3–20, 2006.
- [17] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Math.*, **1**(2):226–251, 2004.
- [18] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures Algorithms*, **6**(2-3):161–179, (1995).
- [19] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.*, **7**(3):295–305, (1998).
- [20] M.E.J. Newman. Assortative mixing in networks. *Physical Review Letters*, **89**(20):208701, (2002).
- [21] M.E.J. Newman. Mixing patterns in networks. *Physical Review E*, **67**(2):026126, (2003).
- [22] M.E.J. Newman. The structure and function of complex networks. *SIAM Rev.*, **45**(2):167–256, 2003.
- [23] M.E.J. Newman. *Networks: an introduction*. Oxford Univsity Press, 2010.
- [24] M.E.J. Newman, S. Strogatz, and D. Watts. Random graphs with arbitrary degree distribution and their application. *Phys. Rev. E*, **64**:026118, 1–17, (2000).
- [25] S. I. Resnick. *Heavy-tail Phenomena*. Springer, New York, 2007.
- [26] C. Spearman. The proof and measurement of association between two things, *The American Journal of Psychology*, **15**(1):72–101, 1904.
- [27] 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*, pages 1113–1114. ACM Press New York, NY, USA, 2008.
- [28] Y. Volkovich, N. Litvak, and B. Zwart. Extremal Dependencies and Rank Correlations in Power Law Networks, In: J. Zhou, O. Akan, P. Bellavista et al. (Eds), *Complex Sciences*, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, **5**:1642–1653, Springer Berlin Heidelberg, 2009.