

ISSN 2220-5438

Reprint from

Moscow Journal

of Combinatorics and Number Theory



Volume 4 • Issue 4

2014

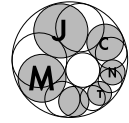
Moscow Journal

of Combinatorics and Number Theory

Volume 4 • Issue 4

2014

URSS



Convergence of rank based degree-degree correlations in random directed networks

Pim van der Hoorn (Enschede), Nelly Litvak (Enschede)

Abstract: We introduce, and analyze, three measures for degree-degree dependencies, also called degree assortativity, in directed random graphs, based on Spearman's rho and Kendall's tau. We prove statistical consistency of these measures in general random graphs and show that the directed Configuration Model can serve as a null model for our degree-degree dependency measures. Based on these results we argue that the measures we introduce should be preferred over Pearson's correlation coefficients, when studying degree-degree dependencies, since the latter has several issues in the case of large networks with scale-free degree distributions.

Keywords: Degree-degree dependencies, rank correlations, directed random graphs, directed configuration model, Spearman's rho, Kendall's tau

AMS Subject classification: 62H20, 05C80

Received: 26.08.2014; **revised:** 30.10.2014

1. Introduction

This paper investigates statistical consistency of rank correlation measures for dependencies between in- and/or out-degrees on both sides of a randomly sampled edge in large directed networks, such as the World Wide Web, Wikipedia, or Twitter. These dependencies, also called the assortativity of the network, degree correlations, or degree-degree dependencies, represent an important topological property of real-

world networks, and they have received a vast attention in the literature, starting with the work of Newman [12, 13].

The underlying question that motivates analysis of degree-degree dependencies is whether nodes of high in- or out-degree are more likely to be connected to nodes of high or low in- or out-degree. These dependencies have been shown to influence many topological features of networks, among others, behavior of epidemic spreading [1], social consensus in Twitter [9], stability of P2P networks under attack [15] and network observability [6]. Therefore, being able to properly measure degree-degree dependencies is essential in modern network analysis.

Given a network, represented by a directed graph, a measurement of degree-degree dependency usually consists of computing some expression that is defined by the degrees at both sides of the edges. Here the value on each edge can be seen as a realization of some unknown ‘true’ parameter that characterizes the degree-degree dependency.

Currently, the most commonly used measure for degree-degree dependencies is a so-called *assortativity coefficient*, introduced in [12, 13], that computes Pearson’s correlation coefficient for the degrees at both sides of an edge. However, this dependency measure suffers from the fact that most real-world networks have highly skewed degree distributions, also called *scale-free* distributions, formally described by power laws, or more formally, regularly varying distributions. Indeed, when the (in- or out-) degree at the end of a random edge has infinite variance, then Pearson’s coefficient is ill-defined. As a result, the dependency measure suggested in [12, 13] depends on the graph size and converges to a non-negative number in the infinite network size limit, as was pointed out in several papers [5, 8]. The detailed mathematical analysis and examples for undirected graphs have been given in [7], and for directed graphs in our recent work [17]. Thus, Pearson’s correlation coefficient is not suitable for measuring degree-degree dependencies in most real-world directed networks.

The fact that the most commonly used degree correlation measure has obvious mathematical flaws, motivates for design and analysis of new estimators. Despite the importance of degree-degree dependencies and vast interest from the research community, this remains a largely open problem.

In [7] it was suggested to use a rank correlation measure, Spearman’s rho, and it was proved that under general regularity conditions, this measure indeed converges to its correct population value. Both configuration model and preferential

attachment model [16] were proved to satisfy these conditions. In [17] we proposed three rank correlation measures, based on Spearman's rho and Kendall's tau, as defined for integer valued random variables, cf. [11], and we compared these measures to Pearson's correlation coefficient on Wikipedia graphs for nine different languages.

In this paper we first prove that, under the convergence assumption of the empirical two-dimensional distributions of the degrees on both sides of a random edge, the rank correlations defined in [17] are indeed statistically consistent estimators of degree-degree dependencies. We obtain their limiting values in terms of the limiting distributions of the degrees.

Next, we apply our results to the recently developed directed Configuration Model [2]. Roughly speaking, in this model, each node is given a random number of in- and out-bound stubs, that are subsequently connected to each other at random. Since multiple edges and self-loops may appear as a result of such random wiring, [2] presents two versions of the directed Configuration Model. The *repeated* version repeats the wiring until the resulting graph is simple, while the *erased* version merges multiple edges and removes self-loops to obtain a simple graph.

We analyze our suggested rank correlation measures in the Repeated and Erased Configuration Model, as described in [2], and prove that all three measures converge to zero in both models. This result is not very surprising for the repeated model, since we connect vertices uniformly at random. However, in the erased scenario, the graph is made simple by design, and this might contribute to the network showing negative degree-degree dependencies as observed and discussed in, for instance, [10, 14]. Our result shows that such negative degree-degree dependencies vanish for sufficiently large graphs, and thus both flavors of the directed Configuration Model can be used as 'null model' for our three rank correlation measures.

By proving consistency of three estimators for degree-degree dependencies in directed networks, and providing an easy-to-construct null model for these estimators, this paper makes an important step towards assessing statistical significance of degree-degree dependencies in a mathematically rigorous way.

This paper is structured as follows. In Section 2 we introduce notations, used throughout this paper. Then, in Section 3, we prove a general theorem concerning statistical consistency of estimators for Spearman's rho and Kendall's tau on integer-valued data. This result is applied in Section 4 in the setting of random graphs to prove the convergence in the infinite size graph limit of the three degree-degree

dependency measures from [17], based on Spearman's rho and Kendall's tau. We analyze both the Repeated and Erased Directed Configuration Model in Section 5.

2. Notations and definitions

Throughout the paper, if X and Y are random variables we denote their distribution functions by F_X and F_Y , respectively, and their joint distribution by $H_{X,Y}$. For integer valued random variables X, Y and $k, l \in \mathbb{Z}$ we will often use the following notations:

$$\mathcal{F}_X(k) = F_X(k) + F_X(k-1), \quad (2.1)$$

$$\begin{aligned} \mathcal{H}_{X,Y}(k, l) &= H_{X,Y}(k, l) + H_{X,Y}(k-1, l) + \\ &\quad + H_{X,Y}(k, l-1) + H_{X,Y}(k-1, l-1). \end{aligned} \quad (2.2)$$

If Z is a random element, we define the function $F_{X|Z} : \mathbb{R} \times \Omega \rightarrow [0, 1]$ by

$$F_{X|Z}(x, \omega) = \mathbb{E}[I\{X \leq x\} | Z](\omega),$$

where $I\{X \leq x\}$ denotes the indicator of the event $\{\omega : X(\omega) \leq x\}$. We furthermore define the random variable $F_{X|Z}(Y)$ by

$$F_{X|Z}(Y)(\omega) = F_{X|Z}(Y(\omega), \omega),$$

and we write $F_{X|Z}(x)$ to indicate the random variable $\mathbb{E}[I\{X \leq x\} | Z]$. With these notations it follows that if X' is an independent copy of X , then

$$\begin{aligned} \mathbb{E}[I\{X' \leq X\} | Z] &= \int_{\mathbb{R}} \int_{\mathbb{R}} I\{z \leq x\} d\mathbb{P}(z|Z) d\mathbb{P}(x|Z) = \\ &= \int_{\mathbb{R}} \mathbb{E}[I\{X' \leq x\} | Z] d\mathbb{P}(x|Z) = \\ &= \mathbb{E}[F_{X|Z}(X) | Z]. \end{aligned}$$

Using similar definitions for $H_{X,Y|Z}(x, y, \omega)$ and $H_{X,Y|Z}(X, Y)$ we get, if (X', Y') and (X'', Y'') are independent copies of (X, Y) , that

$$\mathbb{E}[I\{X' \leq X\} I\{Y'' \leq Y\} | Z] = \mathbb{E}[H_{X,Y|Z}(X, Y) | Z].$$

For integer valued random variables X and Y , the random variables $\mathcal{F}_{X|Z}(k)$ and $\mathcal{H}_{X,Y|Z}(k, l)$ are defined similarly to (2.1) and (2.2), using $F_{X|Z}(k)$ and $H_{X,Y|Z}(k, l)$, respectively.

We introduce the following notion of convergence, related to convergence in distribution.

DEFINITION 1. *Let $\{X_n\}_{n \in \mathbb{N}}$ and X be random variables and $\{Z_n\}_{n \in \mathbb{N}}$ be a sequence of random elements. We say that X_n converges in distribution to X conditioned on Z_n and write*

$$(X_n | Z_n) \Rightarrow X \quad \text{as } n \rightarrow \infty$$

if and only if for all continuous, bounded $h : \mathbb{R} \rightarrow \mathbb{R}$

$$\mathbb{E}[h(X_n) | Z_n] \xrightarrow{\mathbb{P}} \mathbb{E}[h(X)] \quad \text{as } n \rightarrow \infty.$$

Here $\xrightarrow{\mathbb{P}}$ denotes convergence in probability. Note that if h is bounded then $\mathbb{E}[h(X_n) | Z_n]$ is bounded almost everywhere, hence $\lim_{n \rightarrow \infty} \mathbb{E}[h(X_n)] = \lim_{n \rightarrow \infty} \mathbb{E}[\mathbb{E}[h(X_n) | Z_n]] = \mathbb{E}[h(X)]$. Therefore, $(X_n | Z_n) \Rightarrow X$ implies that $X_n \Rightarrow X$, where we write \Rightarrow for convergence in distribution. Similar to convergence in distribution, it holds that Definition 1 is equivalent to

$$F_{X_n | Z_n}(k) \xrightarrow{\mathbb{P}} F_X(k) \quad \text{as } n \rightarrow \infty, \quad \text{for all } k \in \mathbb{Z}.$$

In this paper we use a continuization principle, applied for instance in [11], where we transform given discrete random variables in continuous ones. From here on we will work with integer valued random variables instead of arbitrary discrete random variables.

DEFINITION 2. *Let X be an integer valued random variable and U a uniformly distributed random variable on $[0, 1)$ independent of X . Then we define the continuization of X as*

$$\tilde{X} = X + U.$$

We will refer to U as the *continuous part* of \tilde{X} . We remark that although we have chosen U to be uniform we could instead take any continuous random variable on $[0, 1)$ with strictly increasing cdf, cf. [4].

3. Rank correlations for integer valued random variables

We will use the rank correlations Spearman's rho and Kendall's tau for integer valued random variables as defined in [11]. Below we will state these and rewrite them in terms of the functions \mathcal{F} and \mathcal{H} , defined in (2.1) and (2.2) respectively. We will then proceed, defining estimators for these correlations and prove that, under natural conditions, these converge to the correct value.

3.1. Spearman's rho

Given two integer valued random variables X and Y , Spearman's rho $\rho(X, Y)$ is defined as, c.f. [11]

$$\begin{aligned} \rho(X, Y) = & 3 \left(\mathbb{P}(X < X', Y < Y'') + \mathbb{P}(X \leq X', Y < Y'') + \right. \\ & \left. + \mathbb{P}(X < X', Y \leq Y'') + \mathbb{P}(X \leq X', Y \leq Y'') - 1 \right), \end{aligned}$$

where (X', Y') and (X'', Y'') are independent copies of (X, Y) . We will rewrite this expression, starting with a single term:

$$\begin{aligned} \mathbb{P}(X < X', Y < Y'') &= \mathbb{E}[I\{X < X'\} I\{Y < Y''\}] = \\ &= 1 - \mathbb{E}[I\{X' \leq X\}] - \mathbb{E}[I\{Y'' \leq Y\}] + \\ &\quad + \mathbb{E}[I\{X' \leq X\} I\{Y'' \leq Y\}] = \\ &= 1 - \mathbb{E}[F_X(X)] - \mathbb{E}[F_Y(Y)] + \mathbb{E}[F_X(X)F_Y(Y)]. \end{aligned}$$

If we do the same for the other three terms and use (6.3) we obtain,

$$\rho(X, Y) = 3\mathbb{E}[F_X(X)F_Y(Y)] - 3. \quad (3.1)$$

Since, given two continuous random variables \mathcal{X} and \mathcal{Y} , Spearman's rho is defined as

$$\rho(\mathcal{X}, \mathcal{Y}) = 12\mathbb{E}[F_{\mathcal{X}}(\mathcal{X})F_{\mathcal{Y}}(\mathcal{Y})] - 3,$$

Lemma 7 now implies that

$$\rho(X, Y) = \rho(\tilde{X}, \tilde{Y}). \quad (3.2)$$

3.2. Kendall's tau

For two continuous random variables \mathcal{X} and \mathcal{Y} , Kendall's tau $\tau(\mathcal{X}, \mathcal{Y})$ is defined as

$$\tau(\mathcal{X}, \mathcal{Y}) = 4\mathbb{E}[H_{\mathcal{X}, \mathcal{Y}}(\mathcal{X}, \mathcal{Y})] - 1.$$

Given two discrete random variables X and Y , Kendall's Tau can be written as, c.f. [11] Proposition 2.2,

$$\tau(X, Y) = \mathbb{E}[\mathcal{H}_{X, Y}(X, Y)] - 1. \quad (3.3)$$

Similar to Spearman's rho we obtain, using Lemma 7, that

$$\tau(X, Y) = \tau(\tilde{X}, \tilde{Y}). \quad (3.4)$$

Hence applying the continuization principle from Definition 2 on X and Y preserves both rank correlations. We remark that (3.2) and (3.4) were obtained for arbitrary discrete random variables, using a different approach, in [11].

3.3. Convergence for Spearman's rho and Kendall's tau

Let $\{X_n\}_{n \in \mathbb{N}}$ and $\{Y_n\}_{n \in \mathbb{N}}$ be sequences of integer valued random variables. If $(X_n, Y_n) \Rightarrow (X, Y)$, for some integer valued random variables X and Y , then $\lim_{n \rightarrow \infty} \mathbb{E}[\mathcal{F}_{X_n}(X_n)\mathcal{F}_{Y_n}(Y_n)] = \mathbb{E}[\mathcal{F}_X(X)\mathcal{F}_Y(Y)]$ which implies that $\lim_{n \rightarrow \infty} \rho(X_n, Y_n) = \rho(X, Y)$. The next theorem generalizes this to the setting of the convergence of $(X_n, Y_n|Z_n)$, of Definition 1.

THEOREM 1. *Let $\{X_n\}_{n \in \mathbb{N}}$, $\{Y_n\}_{n \in \mathbb{N}}$ be sequences of integer valued random variables for which there exist a sequence $\{Z_n\}_{n \in \mathbb{N}}$ of random elements and two integer valued random variables X and Y such that*

$$(X_n, Y_n|Z_n) \Rightarrow (X, Y) \quad \text{as } n \rightarrow \infty.$$

Then, as $n \rightarrow \infty$,

- i) $3\mathbb{E}[\mathcal{F}_{X_n|Z_n}(X_n)\mathcal{F}_{Y_n|Z_n}(Y_n)|Z_n] - 3 \xrightarrow{\mathbb{P}} \rho(X, Y)$ and
- ii) $\mathbb{E}[\mathcal{H}_{X_n, Y_n|Z_n}(X_n, Y_n)|Z_n] - 1 \xrightarrow{\mathbb{P}} \tau(X, Y)$.

Moreover, we also have convergence of the expectations:

- iii) $\lim_{n \rightarrow \infty} 3\mathbb{E}[\mathcal{F}_{X_n|Z_n}(X_n)\mathcal{F}_{Y_n|Z_n}(Y_n)] - 3 = \rho(X, Y)$ and

$$\text{iv) } \lim_{n \rightarrow \infty} \mathbb{E}[\mathcal{H}_{X_n, Y_n | Z_n}(X_n, Y_n)] - 1 = \tau(X, Y).$$

PROOF. Observe first that since $(X_n, Y_n | Z_n) \Rightarrow (X, Y)$, it follows that for all $k, l \in \mathbb{Z}$, as $n \rightarrow \infty$,

$$F_{X_n | Z_n}(k) \xrightarrow{\mathbb{P}} F_X(k) \quad (3.5)$$

$$F_{Y_n | Z_n}(l) \xrightarrow{\mathbb{P}} F_Y(l) \quad (3.6)$$

$$H_{X_n, Y_n | Z_n}(k, l) \xrightarrow{\mathbb{P}} H_{X, Y}(k, l). \quad (3.7)$$

Moreover, these convergence hold uniformly, since X and Y are integer valued.

i) Using first (3.1) and then applying Lemma 7 and Proposition 9 we obtain,

$$\begin{aligned} & \left| 3\mathbb{E}[\mathcal{F}_{X_n | Z_n}(X_n)\mathcal{F}_{Y_n | Z_n}(Y_n) | Z_n] - 3 - \rho(X, Y) \right| = \\ & = 3 \left| \mathbb{E}[\mathcal{F}_{X_n | Z_n}(X_n)\mathcal{F}_{Y_n | Z_n}(Y_n) | Z_n] - \mathbb{E}[\mathcal{F}_X(X)\mathcal{F}_Y(Y)] \right| = \\ & = 12 \left| \mathbb{E}\left[F_{\tilde{X}_n | Z_n}(\tilde{X}_n)F_{\tilde{Y}_n | Z_n}(\tilde{Y}_n) \mid Z_n\right] - \mathbb{E}\left[F_{\tilde{X}}(\tilde{X})F_{\tilde{Y}}(\tilde{Y})\right] \right| \leq \\ & \leq 12 \left| \mathbb{E}\left[F_{\tilde{X}_n | Z_n}(\tilde{X}_n)F_{\tilde{Y}_n | Z_n}(\tilde{Y}_n) \mid Z_n\right] - \mathbb{E}\left[F_{\tilde{X}}(\tilde{X}_n)F_{\tilde{Y}}(\tilde{Y}_n) \mid Z_n\right] \right| + \\ & \quad + 12 \left| \mathbb{E}\left[F_{\tilde{X}}(\tilde{X}_n)F_{\tilde{Y}}(\tilde{Y}_n) \mid Z_n\right] - \mathbb{E}\left[F_{\tilde{X}}(\tilde{X})F_{\tilde{Y}}(\tilde{Y})\right] \right| \leq \\ & \leq 12 \sup_{x, y \in \mathbb{R}} \left| F_{\tilde{X}_n | Z_n}(x)F_{\tilde{Y}_n | Z_n}(y) - F_{\tilde{X}}(x)F_{\tilde{Y}}(y) \right| + \end{aligned} \quad (3.8)$$

$$+ 12 \left| \mathbb{E}\left[F_{\tilde{X}}(\tilde{X}_n)F_{\tilde{Y}}(\tilde{Y}_n) \mid Z_n\right] - \mathbb{E}\left[F_{\tilde{X}}(\tilde{X})F_{\tilde{Y}}(\tilde{Y})\right] \right|. \quad (3.9)$$

Because the function $h(x, y) = F_{\tilde{X}}(x)F_{\tilde{Y}}(y)$ is continuous and bounded, (3.9) converges in probability to 0. For (3.8) we observe that

$$\begin{aligned} \left| F_{\tilde{X}_n | Z_n}(x)F_{\tilde{Y}_n | Z_n}(y) - F_{\tilde{X}}(x)F_{\tilde{Y}}(y) \right| & \leq \left| F_{\tilde{X}_n | Z_n}(x)F_{\tilde{Y}_n | Z_n}(y) - F_{\tilde{X}_n | Z_n}(x)F_{\tilde{Y}}(y) \right| + \\ & \quad + \left| F_{\tilde{X}_n | Z_n}(x)F_{\tilde{Y}}(y) - F_{\tilde{X}}(x)F_{\tilde{Y}}(y) \right| \leq \\ & \leq \left| F_{\tilde{Y}_n | Z_n}(y) - F_{\tilde{Y}}(y) \right| + \left| F_{\tilde{X}_n | Z_n}(x) - F_{\tilde{X}}(x) \right|. \end{aligned}$$

It now follows that (3.8) converges in probability to 0, since the convergence (3.5) and (3.6) are uniform.

- ii) Here we again use Lemma 7 and Proposition 9, now combined with (3.3) to obtain,

$$\begin{aligned}
 & \left| \mathbb{E} \left[\mathcal{H}_{X_n, Y_n | Z_n}(X_n, Y_n) \mid Z_n \right] - 1 - \tau(X, Y) \right| = \\
 & = \left| \mathbb{E} \left[\mathcal{H}_{X_n, Y_n | Z_n}(X_n, Y_n) \mid Z_n \right] - \mathbb{E}[\mathcal{H}_{X, Y}(X, Y)] \right| = \\
 & = 4 \left| \mathbb{E} \left[H_{\tilde{X}_n, \tilde{Y}_n | Z_n}(\tilde{X}_n, \tilde{Y}_n) \mid Z_n \right] - \mathbb{E} \left[H_{\tilde{X}, \tilde{Y}}(\tilde{X}, \tilde{Y}) \right] \right| \leq \\
 & \leq 4 \left| \mathbb{E} \left[H_{\tilde{X}_n, \tilde{Y}_n | Z_n}(\tilde{X}_n, \tilde{Y}_n) \mid Z_n \right] - \mathbb{E} \left[H_{\tilde{X}, \tilde{Y}}(\tilde{X}_n, \tilde{Y}_n) \mid Z_n \right] \right| + \\
 & \quad + 4 \left| \mathbb{E} \left[H_{\tilde{X}, \tilde{Y}}(\tilde{X}_n, \tilde{Y}_n) \mid Z_n \right] - \mathbb{E} \left[H_{\tilde{X}, \tilde{Y}}(\tilde{X}, \tilde{Y}) \right] \right| \leq \\
 & \leq 4 \sup_{x, y \in \mathbb{R}} \left| H_{\tilde{X}_n, \tilde{Y}_n | Z_n}(x, y) - H_{\tilde{X}, \tilde{Y}}(x, y) \right| + \\
 & \quad + 4 \left| \mathbb{E} \left[H_{\tilde{X}, \tilde{Y}}(\tilde{X}_n, \tilde{Y}_n) \mid Z_n \right] - \mathbb{E} \left[H_{\tilde{X}, \tilde{Y}}(\tilde{X}, \tilde{Y}) \right] \right|.
 \end{aligned}$$

The former term converges in probability to 0 because (3.7) holds uniformly, and for the latter this holds since $h(x, y) = H_{\tilde{X}, \tilde{Y}}(x, y)$ is continuous and bounded.

Since both $\mathbb{E}[\mathcal{F}_{X_n | Z_n}(X_n) \mathcal{F}_{Y_n | Z_n}(Y_n) \mid Z_n]$ and $\mathbb{E}[\mathcal{H}_{X_n, Y_n | Z_n}(X_n, Y_n) \mid Z_n]$ are bounded a. e. we obtain iii) and iv) directly from i) and ii), respectively. \square

4. Rank correlations for random graphs

We now turn to the setting of rank correlations for degree-degree dependencies in random directed graphs. We will first introduce some terminology concerning random graphs. Then we will recall the rank correlations given in [17] and prove statistical consistency of these measures.

4.1. Random graphs

Given a directed graph $G = (V, E)$, we denote by $(D^+(v), D^-(v))_{v \in V}$ the degree sequence where D^+ denotes the out-degree and D^- the in-degree. We adopt

the convention, introduced in [17], to index the degree type by $\alpha, \beta \in \{+, -\}$. Furthermore, we will use the projections $\pi_*, \pi^* : V^2 \rightarrow V$ to distinguish the source and target of a possible edge. That is, if $(v, w) \in V^2$ then $\pi_*(v, w) = v$ and $\pi^*(v, w) = w$. When both projections are applicable we will use π . For $v, w \in V$ we denote by $E(v, w) = \{e \in E \mid \pi_*e = v, \pi^*e = w\}$ the set of all edges from v to w . For $e \in V^2$, we write $E(e) = E(\pi_*e, \pi^*e)$.

Given a set V of vertices we call a graph $G = (V, E)$ random, if for each $e \in V^2$, $|E(e)|$ is a random variable. Since $I \{e \in E\} = I \{|E(e)| > 0\}$, it follows that the former is also a random variable, cf. [3] for a similar definition of random graphs using edge indicators. Therefore, when we refer to G as a random element it is understood that we refer to the random variables $|E(e)|$, for $e \in V^2$.

When G is a random graph, the number of edges in the graph and the degrees of the nodes are random variables defined by $I \{e \in E\}$ and $|E(e)|$, $e \in V^2$:

$$|E| = \sum_{e \in V^2} I \{e \in E\} |E(e)|,$$

$$D^-(v) = \sum_{w \in V} I \{(w, v) \in E\} |E(w, v)|, \quad v \in V,$$

$$D^+(v) = \sum_{w \in V} I \{(v, w) \in E\} |E(v, w)|, \quad v \in V.$$

Given a random graph $G = (V, E)$ we define a uniformly sampled edge \mathcal{E}_G as a two-dimensional random variable on V^2 such that

$$\mathbb{P}(\mathcal{E}_G = e \mid G) = \frac{|E(e)|}{|E|}.$$

When it is clear which graph we are considering, we will use \mathcal{E} instead of \mathcal{E}_G . Let $\alpha, \beta \in \{+, -\}$, $k, l \in \mathbb{N}$ and π be any of the projections π_* and π^* . Then we define

$$F_G^\alpha(k) = F_{D^\alpha(\pi(\mathcal{E}_G)) \mid G}(k), \tag{4.1}$$

$$H_G^{\alpha, \beta}(k, l) = H_{D^\alpha(\pi_*(\mathcal{E}_G)), D^\beta(\pi^*(\mathcal{E}_G)) \mid G}(k, l). \tag{4.2}$$

These functions are the empirical distribution of $D^\alpha(\pi(\mathcal{E}_G))$ and the joint empirical distribution of $D^\alpha(\pi_*(\mathcal{E}_G))$ and $D^\beta(\pi^*(\mathcal{E}_G))$, respectively, given the random graph G . The functions \mathcal{F}_G^α and $\mathcal{H}_G^{\alpha, \beta}$ are defined in a similar way as (2.1) and (2.2), using (4.1)

and (4.2), respectively. In order to keep notations clear, we will, when considering both projections π_* and π^* , always use α to index the degree type of the sources and β to index the degree type of targets. Moreover, we will often write $D^\alpha \pi \mathcal{E}_G$ instead of $D^\alpha (\pi(\mathcal{E}_G))$.

Now we will introduce Spearman's rho and Kendall's tau on random directed graphs and write them in terms of the functions (4.1) and (4.2). This way we will be in a setting similar to the one of Theorem 1 so that we can utilize this theorem to prove statistical consistency of these rank correlations.

4.2. Spearman's Rho

Spearman's rho measure for degree-degree dependencies in directed graphs, introduced in [17], is in fact Pearson's correlation coefficient computed on the ranks of the degrees rather than their actual values. In our setting, this definition is ambiguous because the data has many ties. For example, if the in-degree of node v is d then we will observe $D^- \pi^* e = d$ for at least d edges $e \in E$, plus there will be many more nodes with the same degree. In [17] we consider two possible ways of resolving ties: by assigning a unique rank to each tied value uniformly at random, and by assigning the same, average, rank to all tied values. We denote the ranks resulting from the random and the average resolution of ties by R and \bar{R} , respectively. Formally, for $\alpha, \beta \in \{+, -\}$, we write:

$$R^\alpha \pi_* e = \sum_{f \in E} I \{ D^\alpha \pi_* f + U_f \geq D^\alpha \pi_* e + U_e \}, \quad (4.3)$$

$$R^\beta \pi^* e = \sum_{f \in E} I \{ D^\beta \pi^* f + W_f \geq D^\beta \pi^* e + W_e \}, \quad (4.4)$$

where U, W are independent $|V|^2$ vectors of independent uniform random variables on $[0, 1)$, and

$$\bar{R}^\alpha \pi e = \frac{1}{2} + \sum_{f \in E} I \{ D^\alpha \pi f > D^\alpha \pi e \} + \frac{1}{2} I \{ D^\alpha \pi f = D^\alpha \pi e \}. \quad (4.5)$$

Then the corresponding two versions of Spearman's rho are defined as follows, cf. [17]:

$$\rho_\alpha^\beta(G) = \frac{12 \sum_{e \in E} R^\alpha \pi_*(e) R^\beta \pi^*(e) - 3|E| (|E| + 1)^2}{|E|^3 - |E|}$$

and

$$\bar{\rho}_\alpha^\beta(G) = \frac{4 \sum_{e \in E} \bar{R}^\alpha \pi_*(e) \bar{R}^\beta \pi^*(e) - |E| (|E| + 1)^2}{\text{Var}_*(\bar{R}^\alpha) \text{Var}^*(\bar{R}^\beta)},$$

where

$$\text{Var}_*(\bar{R}^\alpha) = \sqrt{4 \sum_{e \in E} \bar{R}^\alpha \pi_*(e)^2 - |E| (|E| + 1)^2}$$

and

$$\text{Var}^*(\bar{R}^\beta) = \sqrt{4 \sum_{e \in E} \bar{R}^\beta \pi^*(e)^2 - |E| (|E| + 1)^2}.$$

The next proposition relates the random variables $\rho_\alpha^\beta(G)$ and $\bar{\rho}_\alpha^\beta(G)$ to the random variable

$$\mathbb{E} \left[\mathcal{F}_G^\alpha (D^\alpha \pi_* \mathcal{E}) \mathcal{F}_G^\beta (D^\beta \pi^* \mathcal{E}) \mid G \right]. \quad (4.6)$$

PROPOSITION 1. *Let $G = (V, E)$ be a random graph, \mathcal{E} an edge on G sampled uniformly at random and $\alpha, \beta \in \{+, -\}$. Then*

$$i) \quad \frac{1}{|E|} \sum_{e \in E} \frac{\bar{R}^\alpha \pi_* e}{|E|} \frac{\bar{R}^\beta \pi^* e}{|E|} = \frac{1}{4} \mathbb{E} \left[\mathcal{F}_G^\alpha (D^\alpha \pi_* \mathcal{E}) \mathcal{F}_G^\beta (D^\beta \pi^* \mathcal{E}) \mid G \right] + o_{\mathbb{P}}(|E|^{-1})$$

and

$$ii) \quad \frac{1}{|E|} \sum_{e \in E} \frac{R^\alpha \pi_* e}{|E|} \frac{R^\beta \pi^* e}{|E|} = \frac{1}{4} \mathbb{E} \left[\mathcal{F}_G^\alpha (D^\alpha \pi_* \mathcal{E}) \mathcal{F}_G^\beta (D^\beta \pi^* \mathcal{E}) \mid G \right] + o_{\mathbb{P}}(|E|^{-1}).$$

PROOF.

i) Let \mathcal{E}' be an independent copy of \mathcal{E} and $e \in V^2$. Then it follows from (4.5) that

$$\begin{aligned} \frac{\bar{R}^\alpha \pi e}{|E|} &= \frac{1}{2|E|} + \sum_{f \in E} \frac{1}{|E|} I\{D^\alpha \pi f > D^\alpha \pi e\} + \frac{1}{2|E|} I\{D^\alpha \pi f = D^\alpha \pi e\} = \\ &= 1 + \frac{1}{2|E|} - \frac{1}{2|E|} \sum_{f \in E} I\{D^\alpha \pi f \leq D^\alpha \pi e\} + I\{D^\alpha \pi f \leq D^\alpha \pi e - 1\} = \\ &= 1 + \frac{1}{2|E|} - \frac{1}{2} \sum_{f \in V^2} (I\{D^\alpha \pi f \leq D^\alpha \pi e\} + I\{D^\alpha \pi f \leq D^\alpha \pi e - 1\}) \frac{|E(f)|}{|E|} = \end{aligned}$$

$$\begin{aligned}
&= 1 + \frac{1}{2|E|} - \frac{1}{2} \sum_{f \in V^2} (I\{D^\alpha \pi f \leq D^\alpha \pi e\} + \\
&\quad + I\{D^\alpha \pi f \leq D^\alpha \pi e - 1\}) \mathbb{P}(\mathcal{E}' = f|G) = \\
&= 1 + \frac{1}{2|E|} - \frac{1}{2} (F_G^\alpha(D^\alpha \pi e) + F_G^\alpha(D^\alpha \pi e - 1)) = \\
&= 1 + \frac{1}{2|E|} - \frac{1}{2} \mathcal{F}_G^\alpha(D^\alpha \pi e). \tag{4.7}
\end{aligned}$$

Using a similar expression for $(\bar{R}^\beta \pi_* e) / |E|$ we obtain,

$$\begin{aligned}
&\frac{1}{|E|} \sum_{e \in E} \frac{\bar{R}^\alpha \pi_* e}{|E|} \frac{\bar{R}^\beta \pi_* e}{|E|} = \\
&= \frac{1}{|E|} \sum_{e \in E} \left(1 + \frac{1}{2|E|} - \frac{1}{2} \mathcal{F}_G^\alpha(D^\alpha \pi_* e) \right) \left(1 + \frac{1}{2|E|} - \frac{1}{2} \mathcal{F}_G^\beta(D^\beta \pi_* e) \right) = \\
&= \mathbb{E} \left[\left(1 + \frac{1}{2|E|} - \frac{1}{2} \mathcal{F}_G^\alpha(D^\alpha \pi_* \mathcal{E}) \right) \left(1 + \frac{1}{2|E|} - \frac{1}{2} \mathcal{F}_G^\beta(D^\beta \pi_* \mathcal{E}) \right) \middle| G \right].
\end{aligned}$$

Rearranging the terms yields

$$\begin{aligned}
&\frac{1}{|E|} \sum_{e \in E} \frac{\bar{R}^\alpha \pi_* e}{|E|} \frac{\bar{R}^\beta \pi_* e}{|E|} = \frac{1}{4} \mathbb{E} \left[\mathcal{F}_G^\alpha(D^\alpha \pi_* \mathcal{E}) \mathcal{F}_G^\beta(D^\beta \pi_* \mathcal{E}) \middle| G \right] + \\
&\quad + 1 - \frac{1}{2} \mathbb{E} \left[\mathcal{F}_G^\alpha(D^\alpha \pi_* \mathcal{E}) + \mathcal{F}_G^\beta(D^\beta \pi_* \mathcal{E}) \middle| G \right] + o_{\mathbb{P}}(|E|^{-1}). \tag{4.8}
\end{aligned}$$

Since the sum over all average ranks equals $|E|(|E| + 1)/2$, it follows that

$$\frac{1}{2} + \frac{1}{2|E|} = \frac{1}{|E|} \sum_{e \in E} \frac{\bar{R}^\alpha \pi e}{|E|} = 1 + \frac{1}{2|E|} - \frac{1}{2} \mathbb{E}[\mathcal{F}_G^\alpha(D^\alpha \pi e) | G],$$

from which we deduce that

$$\mathbb{E}[\mathcal{F}_G^\alpha(D^\alpha \pi e) | G] = 1. \tag{4.9}$$

The result now follows by inserting (4.9) in (4.8).

ii) Again, let \mathcal{E}' be an independent copy of \mathcal{E} and $\alpha, \beta \in \{+, -\}$. For $x, y \in \mathbb{R}$, we write $\widetilde{F}_G^\alpha(x) = F_{\widetilde{D^\alpha \pi_* \mathcal{E}}|G}(x)$ and similarly $\widetilde{F}_G^\beta(y) = F_{\widetilde{D^\beta \pi_* \mathcal{E}}|G}(y)$. Then we have,

$$\begin{aligned} \frac{R^\alpha \pi_* e}{|E|} &= \frac{1}{|E|} \sum_{f \in E} I \{ D^\alpha \pi_* f + U_f \geq D^\alpha \pi_* e + U_e \} = \\ &= \frac{1}{|E|} \sum_{f \in E} I \{ D^\alpha \pi_* f + U_f > D^\alpha \pi_* e + U_e \} + I \{ f = e \} = \\ &= 1 - \mathbb{E} [I \{ D^\alpha \pi_* \mathcal{E}' + U_{\mathcal{E}'} \leq D^\alpha \pi_* e + U_e \} | G] + \frac{1}{|E|} = \\ &= 1 - \widetilde{F}_G^\alpha (D^\alpha \pi_* e + U_e) + \frac{1}{|E|}. \end{aligned} \quad (4.10)$$

Using similar calculations we get

$$\frac{R^\beta \pi^* e}{|E|} = 1 - \widetilde{F}_G^\beta (D^\beta \pi^* e + W_e) + \frac{1}{|E|}. \quad (4.11)$$

Now, using both (4.10) and (4.11), we obtain,

$$\begin{aligned} \frac{1}{|E|} \sum_{e \in E} \frac{R^\alpha \pi_* e}{|E|} \frac{R^\beta \pi^* e}{|E|} &= 1 + \frac{2}{|E|} + \frac{1}{|E|^2} + \\ &+ \frac{1}{|E|} \sum_{e \in E} \widetilde{F}_G^\alpha (D^\alpha \pi_* e + U_e) \widetilde{F}_G^\beta (D^\beta \pi^* e + W_e) + \\ &- \left(1 + \frac{1}{|E|} \right) \frac{1}{|E|} \sum_{e \in E} \left(\widetilde{F}_G^\alpha (D^\alpha \pi_* e + U_e) + \widetilde{F}_G^\beta (D^\beta \pi^* e + W_e) \right) = \\ &= 1 + \frac{2}{|E|} + \frac{1}{|E|^2} + \mathbb{E} \left[\widetilde{F}_G^\alpha (\widetilde{D^\alpha \pi_* \mathcal{E}}) \widetilde{F}_G^\beta (\widetilde{D^\beta \pi^* \mathcal{E}}) | G \right] + \\ &- \left(1 + \frac{1}{|E|} \right) \left(\mathbb{E} \left[\widetilde{F}_G^\alpha (\widetilde{D^\alpha \pi_* \mathcal{E}}) | G \right] + \mathbb{E} \left[\widetilde{F}_G^\beta (\widetilde{D^\beta \pi^* \mathcal{E}}) | G \right] \right) = \\ &= \frac{1}{4} \mathbb{E} \left[\mathcal{F}_G^\alpha (D^\alpha \pi_* \mathcal{E}) \mathcal{F}_G^\beta (D^\beta \pi^* \mathcal{E}) | G \right] + \frac{1}{|E|} + \frac{1}{|E|^2}. \end{aligned}$$

The last line follows by first using Propositions 8 and 9 to rewrite the conditional expectations and then applying (4.9). \square

4.3. Kendall's Tau

The definition for $\tau_\alpha^\beta(G)$ is, cf. [17],

$$\tau_\alpha^\beta(G) = \frac{2(\mathcal{N}_C(G) - \mathcal{N}_D(G))}{|E|(|E| - 1)},$$

where $\mathcal{N}_C(G)$ and $\mathcal{N}_D(G)$ denote the number of concordant and discordant pairs, respectively, among $(D^\alpha \pi_* e, D^\beta \pi^* e)_{e \in E}$. We recall that a pair $(D^\alpha \pi_* e, D^\beta \pi^* e)$ and $(D^\alpha \pi_* f, D^\beta \pi^* f)$, for $e, f \in E$ is called (discordant) concordant if

$$(D^\alpha \pi_* e - D^\alpha \pi_* f) (D^\beta \pi^* e - D^\beta \pi^* f) (< 0) > 0.$$

Therefore we have, for the concordant pairs,

$$\begin{aligned} \frac{2}{|E|^2} \mathcal{N}_C(G) &= \frac{1}{|E|^2} \sum_{e, f \in E} I \left\{ D^\alpha \pi_*(f) < D^\alpha \pi_*(e), D^\beta \pi^*(f) < D^\beta \pi^*(e) \right\} + \\ &+ \frac{1}{|E|^2} \sum_{e, f \in E} I \left\{ D^\alpha \pi_*(f) > D^\alpha \pi_*(e), D^\beta \pi^*(f) > D^\beta \pi^*(e) \right\} = \\ &= \mathbb{E} \left[H_G^{\alpha, \beta} \left(D^\alpha \pi_* \mathcal{E} - 1, D^\beta \pi^* \mathcal{E} - 1 \right) \middle| G \right] + \\ &+ 1 - \mathbb{E} \left[F_G^\alpha \left(D^\alpha \pi_* \mathcal{E} \right) \middle| G \right] - \mathbb{E} \left[F_G^\beta \left(D^\beta \pi^* \mathcal{E} \right) \middle| G \right] + \\ &+ \mathbb{E} \left[H_G^{\alpha, \beta} \left(D^\alpha \pi_* \mathcal{E}, D^\beta \pi^* \mathcal{E} \right) \middle| G \right]. \end{aligned}$$

In a similar fashion we get for the discordant pairs

$$\begin{aligned} \frac{2}{|E|^2} \mathcal{N}_D(G) &= \mathbb{E} \left[F_G^\alpha \left(D^\alpha \pi_* \mathcal{E} - 1 \right) \middle| G \right] + \mathbb{E} \left[F_G^\beta \left(D^\beta \pi^* \mathcal{E} - 1 \right) \middle| G \right] + \\ &- \mathbb{E} \left[H_G^{\alpha, \beta} \left(D^\alpha \pi_* \mathcal{E} - 1, D^\beta \pi^* \mathcal{E} \right) \middle| G \right] + \\ &- \mathbb{E} \left[H_G^{\alpha, \beta} \left(D^\alpha \pi_* \mathcal{E}, D^\beta \pi^* \mathcal{E} - 1 \right) \middle| G \right]. \end{aligned}$$

Combining the above with (4.9) we conclude that

$$\tau_\alpha^\beta(G) = \mathbb{E} \left[H_G^{\alpha, \beta} \left(D^\alpha \pi_* \mathcal{E}, D^\beta \pi^* \mathcal{E} \right) \middle| G \right] - 1 + o_{\mathbb{P}}(|E|^{-1}). \quad (4.12)$$

4.4. Statistical consistency of rank correlations

We will now prove that the rank correlations defined in the previous two sections are, under natural regularity conditions on the degree sequences, consistent statistical estimators.

For a sequence $\{G_n\}_{n \in \mathbb{N}}$ of random graphs with $|V_n| = n$, it is common in the theory of random graphs to assume convergence of the empirical degree distributions, see for instance Condition 7.5 in [16], Condition 4.1 in [2]. Here, similarly to [7], we impose the following regularity condition on the degrees at the end points of edges.

Conclusion 1. *Given a sequence $\{G_n\}_{n \in \mathbb{N}}$ of random graphs with $|V_n| = n$ and $\alpha, \beta \in \{+, -\}$ there exist integer valued random variables \mathcal{D}^α and \mathcal{D}^β , not concentrated in a single point, such that*

$$\left(D_n^\alpha \pi_* \mathcal{E}_n, D_n^\beta \pi^* \mathcal{E}_n \mid G_n \right) \Rightarrow \left(\mathcal{D}^\alpha, \mathcal{D}^\beta \right) \quad \text{as } n \rightarrow \infty,$$

where \mathcal{E}_n is a uniformly sampled edge in G_n .

In the previous two sections it was shown that $\rho_\alpha^\beta(G), \bar{\rho}_\alpha^\beta(G)$ and $\tau_\alpha^\beta(G)$ on a random graph G are related to, respectively,

$$\mathbb{E} \left[\mathcal{F}_G^\alpha \left(D^\alpha \pi_* \mathcal{E} \right) \mathcal{F}_G^\beta \left(D^\beta \pi^* \mathcal{E} \right) \mid G \right] \quad \text{and} \quad \mathbb{E} \left[\mathcal{H}_G^{\alpha, \beta} \left(D^\alpha \pi_* \mathcal{E}, D^\beta \pi^* \mathcal{E} \right) \mid G \right].$$

Note that these are in fact empirical versions of the functions appearing in the definitions of Spearman’s rho and Kendall’s tau, cf. (3.1) and (3.3). The following result formalizes these observations and states that under Condition 1, $\rho_\alpha^\beta(G_n), \bar{\rho}_\alpha^\beta(G_n)$ and $\tau_\alpha^\beta(G_n)$ are indeed consistent statistical estimators of correlation measures associated with Spearman’s rho and Kendall’s tau.

THEOREM 2. *Let $\alpha, \beta \in \{+, -\}$ and $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of graphs satisfying Condition 1 such that as $n \rightarrow \infty, |E_n| \xrightarrow{\mathbb{P}} \infty$. Then, as $n \rightarrow \infty,$*

- i) $\rho_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} \rho \left(\mathcal{D}^\alpha, \mathcal{D}^\beta \right),$
- ii) $\bar{\rho}_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} \frac{\rho \left(\mathcal{D}^\alpha, \mathcal{D}^\beta \right)}{3 \sqrt{S_{\mathcal{D}^\alpha} \left(\mathcal{D}^\alpha \right) S_{\mathcal{D}^\beta} \left(\mathcal{D}^\beta \right)}},$

where $S_{\mathcal{D}^\alpha} \left(\mathcal{D}^\alpha \right) = \mathbb{E} \left[F_{\mathcal{D}^\alpha} \left(\mathcal{D}^\alpha \right) F_{\mathcal{D}^\alpha} \left(\mathcal{D}^\alpha - 1 \right) \right]$

and

$$\text{iii) } \tau_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} \tau(\mathcal{D}^\alpha, \mathcal{D}^\beta).$$

Moreover, we have convergence of the first moments:

$$\text{iv) } \lim_{n \rightarrow \infty} \mathbb{E} \left[\rho_\alpha^\beta(G_n) \right] = \rho(\mathcal{D}^\alpha, \mathcal{D}^\beta),$$

$$\text{v) } \lim_{n \rightarrow \infty} \mathbb{E} \left[\bar{\rho}_\alpha^\beta(G_n) \right] = \frac{\rho(\mathcal{D}^\alpha, \mathcal{D}^\beta)}{3\sqrt{S_{\mathcal{D}^\alpha}(\mathcal{D}^\alpha) S_{\mathcal{D}^\beta}(\mathcal{D}^\beta)}}$$

and

$$\text{vi) } \lim_{n \rightarrow \infty} \mathbb{E} \left[\tau_\alpha^\beta(G_n) \right] = \tau(\mathcal{D}^\alpha, \mathcal{D}^\beta).$$

PROOF.

i) By Proposition 1 we have that

$$\begin{aligned} \frac{12}{|E_n|} \sum_{e \in E_n} \frac{R_n^\alpha \pi_* e}{|E_n|} \frac{R_n^\beta \pi^* e}{|E_n|} &= 3\mathbb{E} \left[\mathcal{F}_{G_n}^\alpha(D_n^\alpha \pi_* \mathcal{E}_n) \mathcal{F}_{G_n}^\beta(D_n^\beta \pi^* \mathcal{E}_n) \mid G_n \right] + \\ &+ o_{\mathbb{P}}(|E_n|^{-1}). \end{aligned}$$

From this and the fact that $|E_n| \xrightarrow{\mathbb{P}} \infty$ it follows that,

$$\begin{aligned} \rho_\alpha^\beta(G_n) &= \frac{1}{1 - |E_n|^{-2}} \left(\frac{12}{|E_n|} \sum_{e \in E_n} \frac{R_n^\alpha \pi_* e}{|E_n|} \frac{R_n^\beta \pi^* e}{|E_n|} - 3 \frac{|E_n| (|E_n| + 1)^2}{|E_n|^3} \right) = \\ &= 3\mathbb{E} \left[\mathcal{F}_{G_n}^\alpha(D_n^\alpha \pi_* \mathcal{E}_n) \mathcal{F}_{G_n}^\beta(D_n^\beta \pi^* \mathcal{E}_n) \mid G_n \right] - 3 + o_{\mathbb{P}}(|E_n|^{-1}) \xrightarrow{\mathbb{P}} \\ &\xrightarrow{\mathbb{P}} \rho(\mathcal{D}^\alpha, \mathcal{D}^\beta) \quad \text{as } n \rightarrow \infty, \end{aligned}$$

where the last line follows from Theorem 1.

ii) From (4.7) it follows that,

$$\left(\frac{\bar{R}_n^\alpha \pi e}{|E_n|} \right)^2 = \left(1 + \frac{1}{2|E_n|} \right)^2 - \left(1 + \frac{1}{2|E_n|} \right) \mathcal{F}_{G_n}^\alpha(D^\alpha \pi e) + \frac{1}{4} \mathcal{F}_{G_n}^\alpha(D^\alpha \pi e)^2.$$

Therefore,

$$\frac{1}{|E_n|} \sum_{e \in E_n} \left(\frac{\bar{R}_n^\alpha \pi e}{|E_n|} \right)^2 = \left(1 + \frac{1}{2|E_n|} \right)^2 + \frac{1}{4} \mathbb{E} \left[\mathcal{F}_{G_n}^\alpha(D^\alpha \pi \mathcal{E}_n)^2 \mid G_n \right] +$$

$$\begin{aligned}
 & - \left(1 + \frac{1}{2|E_n|} \right) \mathbb{E} \left[\mathcal{F}_{G_n}^\alpha (D^\alpha \pi \mathcal{E}_n) \mid G_n \right] = \\
 & = 1 + \frac{1}{4} \mathbb{E} \left[\mathcal{F}_{G_n}^\alpha (D^\alpha \pi \mathcal{E}_n)^2 \mid G_n \right] + \\
 & \quad - \mathbb{E} \left[\mathcal{F}_{G_n}^\alpha (D^\alpha \pi \mathcal{E}_n) \mid G_n \right] + o_{\mathbb{P}}(|E_n|^{-1}) \xrightarrow{\mathbb{P}} \\
 & \xrightarrow{\mathbb{P}} 1 + \frac{1}{4} \mathbb{E} \left[\mathcal{F}_{\mathcal{D}^\alpha} (D^\alpha)^2 \right] - \mathbb{E} \left[\mathcal{F}_{\mathcal{D}^\alpha} (D^\alpha) \right] \quad \text{as } n \rightarrow \infty = \\
 & = \frac{1}{4} + \frac{1}{4} \mathbb{E} \left[F_{\mathcal{D}^\alpha} (D^\alpha) F_{\mathcal{D}^\alpha} (D^\alpha - 1) \right],
 \end{aligned}$$

where we used Lemma 6 for the last line. It follows that, as $n \rightarrow \infty$,

$$\frac{4}{|E_n|} \sum_{e \in E_n} \left(\frac{\bar{R}_n^\alpha \pi e}{|E_n|} \right)^2 - \frac{|E_n| (|E_n| + 1)^2}{|E_n|^3} \xrightarrow{\mathbb{P}} \mathbb{E} \left[F^\alpha (D^\alpha) F^\alpha (D^\alpha - 1) \right].$$

Since \mathcal{D}^α and \mathcal{D}^β are not concentrated in one point the above term is non-zero. Now, combining this with Proposition 1 i) and applying Theorem 1, we obtain

$$\bar{\rho}_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} \frac{\rho(\mathcal{D}^\alpha, \mathcal{D}^\beta)}{3 \sqrt{S_{\mathcal{D}^\alpha}(\mathcal{D}^\alpha) S_{\mathcal{D}^\beta}(\mathcal{D}^\beta)}} \quad \text{as } n \rightarrow \infty.$$

iii) Combining (4.12) with Theorem 1 yields, as $n \rightarrow \infty$,

$$\begin{aligned}
 \tau_\alpha^\beta(G_n) & = \\
 & = \mathbb{E} \left[\mathcal{H}_{G_n}^{\alpha, \beta} \left(D_n^\alpha \pi_* \mathcal{E}_n, D_n^\beta \pi^* \mathcal{E}_n \right) \mid G_n \right] - 1 + o_{\mathbb{P}}(|E_n|^{-1}) \xrightarrow{\mathbb{P}} \tau(\mathcal{D}^\alpha, \mathcal{D}^\beta).
 \end{aligned} \tag{4.13}$$

Finally, iv),v),vi) now follow from, respectively, i), ii) and iii) since $\rho_\alpha^\beta(G_n)$, $\bar{\rho}_\alpha^\beta(G_n)$ and $\tau_\alpha^\beta(G_n)$ are bounded. □

Comparing results i) and iv) to ii) and v), note that the way in which ties are resolved influences the measure estimated by Spearman’s rho on random directed graphs. In particular, resolving ties uniformly at random yields the value corresponding to Spearman’s rho for the two limiting integer valued random variables \mathcal{D}^α and \mathcal{D}^β as defined in [11], in the infinite size network limit.

5. Directed Configuration Model

In this section we will analyze degree-degree dependencies for the directed *Configuration Model* (CM), as described and analyzed in [2]. First, in Section 5.1, we analyze the model where in- and out-links are connected at random, which, in general, results in a multi-graph. Then we move on to two other models that produce simple graphs: the *Repeated* and *Erased* Configuration Model (RCM and ECM). By applying Theorem 2, in Sections 5.2 and 5.3, we will show that RCM and ECM can be used as null models for the rank correlations ρ , $\bar{\rho}$ and τ .

5.1. General model: multi-graphs

The directed Configuration Model in [2] starts with picking two target distributions F_- , F_+ for the in- and out-degrees, respectively, stochastically bounded from above by regularly varying distributions. We will adopt notations from [2] and let γ and ξ denote random variables with distributions F_- and F_+ , respectively. It is assumed that $\mathbb{E}[\gamma] = \mathbb{E}[\xi] < \infty$. The next step is generating a bi-degree sequence of inbound and outbound stubs. This is done by first taking two independent sequences of n independent copies of γ and ξ , which are then modified into a sequence of in- and outbound stubs

$$\widehat{\mathcal{D}}(G) = \left(\widehat{D}^+(v), \widehat{D}^-(v) \right)_{v \in V},$$

using the algorithm in [2], Section 2.1. This algorithm ensures that the total number of in- and outbound stubs is the same, $|\widehat{E}| = \sum_{v \in V} \widehat{D}^\alpha(v)$, $\alpha \in \{+, -\}$. Using this bi-degree sequence, a graph is build by randomly pairing the stubs to form edges. We call a graph generated by this model a *Configuration Model graph*, or CM graph for short. We remark that a CM graph in general does not need to be simple.

Given a vertex set V , a bi-degree sequence $\widehat{\mathcal{D}}(G)$ and $v \in V$, we denote by v_i^+ , v_j^- for $1 \leq i \leq \widehat{D}^+(v)$ and $1 \leq j \leq \widehat{D}^-(v)$, respectively, the outbound and inbound stubs of v . For $v, w \in V$, we denote by $\{v_i^+ \rightarrow w_j^-\}$ the event that the outbound stub v_i^+ is connected to the inbound stub w_j^- and by $\{v_i^+ \rightarrow w\}$ the event that v_i^+ is connected to an inbound stub of w . By definition of CM, it follows that $\mathbb{P} \left(v_i^+ \rightarrow w_j^- | \widehat{\mathcal{D}}(G) \right) = 1/|\widehat{E}|$ and hence $\mathbb{P} \left(v_i^+ \rightarrow w | \widehat{\mathcal{D}}(G) \right) = \widehat{D}^-(w)/|\widehat{E}|$.

Furthermore we observe that $|\widehat{E}_n(e)| = \sum_{i=1}^{\widehat{D}_n^+ \pi_* e} I \{ (\pi_* e)_i^+ \rightarrow \pi^* e \}$. Given a random

graph G , we denote

$$I_e^{\alpha,\beta}(k, l) = I \{ D^\alpha \pi_* e = k \} I \{ D^\beta \pi^* e = l \},$$

where $\alpha, \beta \in \{+, -\}$, $k, l \in \mathbb{N}$ and $e \in V^2$.

For proper reference we summarize some results from Proposition 2.5, in [2], which we will use in the remainder of this paper.

PROPOSITION 2 ([2], PROPOSITION 2.5). *Let $\widehat{\mathfrak{D}}(G_n)$ be the bi-degree sequence on n vertices, as generated in Section 2.1 of [2], and $k, l \in \mathbb{N}$. Then, as $n \rightarrow \infty$,*

$$\begin{aligned} & \frac{1}{n} \sum_{v \in V_n} I \{ \widehat{D}_n^+ v = k \} I \{ \widehat{D}_n^- v = l \} \xrightarrow{\mathbb{P}} \mathbb{P}(\xi = k) \mathbb{P}(\gamma = l), \\ & \frac{1}{n} \sum_{v \in V_n} \widehat{D}_n^+ v \xrightarrow{\mathbb{P}} \mathbb{E}[\xi] \quad \text{and} \quad \frac{1}{n} \sum_{v \in V_n} \widehat{D}_n^- v \xrightarrow{\mathbb{P}} \mathbb{E}[\gamma]. \end{aligned}$$

Given a random graph $G = (V, E)$, we will use $\mathfrak{D}(G)$ as a short hand notation for its degree sequence $(D^-(v), D^+(v))_{v \in V}$. We emphasize that for a graph generated using an initial bi-degree sequence, the eventual degree sequence $\mathfrak{D}(G)$ can be different from $\widehat{\mathfrak{D}}(G)$. This, for example, is true for the ECM, Section 5.3, where, after the random pairing of the stubs, self-loops are removed and multiple edges are merged.

In order to apply Theorem 2 to a sequence of (multi-)graphs $\{G_n\}_{n \in \mathbb{N}}$ generated by CM, we need to prove that

$$\left(D_n^\alpha \pi_* \mathcal{E}_n, D_n^\beta \pi^* \mathcal{E}_n \mid G_n \right) \Rightarrow \left(\mathcal{D}^\alpha, \mathcal{D}^\beta \right),$$

for some integer valued random variables \mathcal{D}^α and \mathcal{D}^β . For this, it suffices to show that, as $n \rightarrow \infty$,

$$H_{G_n}^{\alpha,\beta}(k, l) \xrightarrow{\mathbb{P}} H_{\mathcal{D}^\alpha, \mathcal{D}^\beta}(k, l),$$

for all $k, l \in \mathbb{N}$. We will prove this by showing that

$$\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n \right] \xrightarrow{\mathbb{P}} \mathbb{P} \left(\mathcal{D}^\alpha = k, \mathcal{D}^\beta = l \right),$$

as $n \rightarrow \infty$, using a second moment argument as follows. Given a sequence $\{G_n\}_{n \in \mathbb{N}}$ of graphs, $\alpha, \beta \in \{+, -\}$ and $k, l \in \mathbb{N}$, we will show that the empirical joint

probability $\mathbb{E}\left[\mathbb{E}\left[I_{\mathcal{E}_n}^{\alpha,\beta}(k,l)\mid G_n\right]\right]$ converges to $\mathbb{P}(D^\alpha = k, D^\beta = l)$. Then we will prove that the variance of $\mathbb{E}\left[I_{\mathcal{E}_n}^{\alpha,\beta}(k,l)\mid G_n\right]$ converges to zero.

We start with expressing the first and second moment of $\mathbb{E}\left[I_{\mathcal{E}_n}^{\alpha,\beta}(k,l)\mid G_n\right]$, for CM graphs, conditioned on the bi-degree sequence $\widehat{\mathcal{D}}(G_n)$ in terms of the degrees. We observe that, for $\alpha, \beta \in \{+, -\}$, $e \in V_n^2$ and $k, l \in \mathbb{N}$, the events $\{D_n^\alpha \pi_* e = k\}$ and $\{D_n^\beta \pi^* e = l\}$ are completely defined by $\widehat{\mathcal{D}}(G_n)$, hence so is $I_e^{\alpha,\beta}(k,l)$. We remark that, since CM leaves the number of inbound and outbound stubs intact, we have $\mathcal{D}(G_n) = \widehat{\mathcal{D}}(G_n)$. However, in this section we will keep using hats, e. g. \widehat{D}_n instead of D_n , to emphasize that G_n can be a multi-graph.

LEMMA 1. *Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of CM graphs with $|V_n| = n$ and $\alpha, \beta \in \{+, -\}$. Then, for each $k, l \in \mathbb{N}$,*

$$i) \mathbb{E}\left[\mathbb{E}\left[I_{\mathcal{E}_n}^{\alpha,\beta}(k,l)\mid G_n\right]\mid \widehat{\mathcal{D}}(G_n)\right] = \sum_{e \in V_n^2} I_e^{\alpha,\beta}(k,l) \frac{\widehat{D}_n^+ \pi_* e \widehat{D}_n^- \pi^* e}{|\widehat{E}_n|^2}$$

and

$$ii) \mathbb{E}\left[\mathbb{E}\left[I_{\mathcal{E}_n}^{\alpha,\beta}(k,l)\mid G_n\right]^2\mid \widehat{\mathcal{D}}(G_n)\right] = \left(\sum_{e \in V_n^2} I_e^{\alpha,\beta}(k,l) \frac{\widehat{D}_n^+ \pi_* e \widehat{D}_n^- \pi^* e}{|\widehat{E}_n|^2}\right)^2 + o_{\mathbb{P}}(1).$$

PROOF.

i)

$$\begin{aligned} \mathbb{E}\left[\mathbb{E}\left[I_{\mathcal{E}_n}^{\alpha,\beta}(k,l)\mid G_n\right]\mid \widehat{\mathcal{D}}(G_n)\right] &= \mathbb{E}\left[\sum_{e \in V_n^2} I_e^{\alpha,\beta}(k,l) \frac{|\widehat{E}_n(e)|}{|\widehat{E}_n|}\mid \widehat{\mathcal{D}}(G_n)\right] = \\ &= \frac{1}{|\widehat{E}_n|} \sum_{e \in V_n^2} I_e^{\alpha,\beta}(k,l) \mathbb{E}\left[|\widehat{E}_n(e)|\mid \widehat{\mathcal{D}}(G_n)\right] = \\ &= \frac{1}{|\widehat{E}_n|} \sum_{e \in V_n^2} I_e^{\alpha,\beta}(k,l) \mathbb{E}\left[\sum_{i=1}^{\widehat{D}_n^+ \pi_* e} I\{(\pi_* e)_i^+ \rightarrow \pi^* e\}\mid \widehat{\mathcal{D}}(G_n)\right] = \\ &= \sum_{e \in V_n^2} I_e^{\alpha,\beta}(k,l) \frac{(\widehat{D}_n^+ \pi_* e) (\widehat{D}_n^- \pi^* e)}{|\widehat{E}_n|^2}. \end{aligned} \tag{5.1}$$

ii) Following similar calculations as above we get,

$$\begin{aligned}
 & \mathbb{E} \left[\mathbb{E} \left[I_{\xi_n}^{\alpha, \beta}(k, l) \mid G_n \right]^2 \mid \widehat{\mathcal{D}}(G_n) \right] = \\
 & = \mathbb{E} \left[\sum_{e, f \in V_n^2} I_e^{\alpha, \beta}(k, l) I_f^{\alpha, \beta}(k, l) \frac{|\widehat{E}_n(e)| |\widehat{E}_n(f)|}{|\widehat{E}_n|^2} \mid \widehat{\mathcal{D}}(G_n) \right] = \tag{5.2} \\
 & = \frac{1}{|\widehat{E}_n|^2} \sum_{e, f \in V_n^2} \left(I_e^{\alpha, \beta}(k, l) I_f^{\alpha, \beta}(k, l) \times \right. \\
 & \quad \left. \sum_{i=1}^{\widehat{D}_n^+ \pi_* e} \sum_{s=1}^{\widehat{D}_n^+ \pi_* f} \mathbb{E} \left[I \{ (\pi_* e)_i^+ \rightarrow \pi^* e \} I \{ (\pi_* f)_s^+ \rightarrow \pi^* f \} \mid \widehat{\mathcal{D}}(G_n) \right] \right). \tag{5.3}
 \end{aligned}$$

We will, for $e, f \in V_n^2$, analyze

$$\frac{1}{|\widehat{E}_n|^2} \sum_{i=1}^{\widehat{D}_n^+ \pi_* e} \sum_{s=1}^{\widehat{D}_n^+ \pi_* f} \mathbb{E} \left[I \{ (\pi_* e)_i^+ \rightarrow \pi^* e \} I \{ (\pi_* f)_s^+ \rightarrow \pi^* f \} \mid \widehat{\mathcal{D}}(G_n) \right] \tag{5.4}$$

for all different cases, $e = f$, $e \cap f = \emptyset$, $e_* = f_*$ and $e^* = f^*$. First, suppose that $e = f$. Then (5.4) equals

$$\frac{1}{|\widehat{E}_n|^2} \sum_{i, s=1}^{\widehat{D}_n^+ \pi_* e} \sum_{j, t=1}^{\widehat{D}_n^+ \pi_* e} \frac{I \{ i = s \} I \{ j = t \}}{|\widehat{E}_n|} + \frac{I \{ i \neq s \} I \{ j \neq t \}}{|\widehat{E}_n| (|\widehat{E}_n| - 1)}.$$

Writing out the sums and using that $e = f$ we obtain,

$$(5.4) = \frac{\widehat{D}_n^+ \pi_* e \widehat{D}_n^- \pi^* e \widehat{D}_n^+ \pi_* f \widehat{D}_n^- \pi^* f}{|\widehat{E}_n|^3 (|\widehat{E}_n| - 1)} + \tag{5.5}$$

$$+ \frac{(\widehat{D}_n^+ \pi_* e) (\widehat{D}_n^- \pi^* e)}{|\widehat{E}_n|^3} + \frac{(\widehat{D}_n^+ \pi_* e) (\widehat{D}_n^- \pi^* e)}{|\widehat{E}_n|^3 (|\widehat{E}_n| - 1)} + \tag{5.6}$$

$$- \frac{(\widehat{D}_n^- \pi^* e)^2 (\widehat{D}_n^+ \pi_* e)}{|\widehat{E}_n|^3 (|\widehat{E}_n| - 1)} - \frac{(\widehat{D}_n^+ \pi_* e)^2 (\widehat{D}_n^- \pi^* e)}{|\widehat{E}_n|^3 (|\widehat{E}_n| - 1)}. \tag{5.7}$$

Since for all $k \geq 0$ and $\kappa \in \{+, -\}$ it holds that

$$\frac{1}{|\widehat{E}_n|^{k+1}} \sum_{v \in V_n} \left(\widehat{D}_n^\kappa v \right)^k \leq \frac{1}{|\widehat{E}_n|^{k+1}} \left(\sum_{v \in V_n} \widehat{D}_n^\kappa v \right)^k = \frac{1}{|\widehat{E}_n|},$$

we deduce that the terms in (5.6) and (5.7) contribute as $o_{\mathbb{P}}(1)$ in (5.3), from which the result for $e = f$ follows. The calculations for the other three cases for $e, f \in V_n^2$ are similar and are hence omitted. \square

As a direct consequence we have the following

PROPOSITION 3. *Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of CM graphs with $|V_n| = n$ and $\alpha, \beta \in \{+, -\}$. Then, for each $k, l \in \mathbb{N}$, as $n \rightarrow \infty$,*

$$\left| \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right]^2 \mid \widehat{\mathcal{D}}(G_n) \right] - \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right]^2 \right| \xrightarrow{\mathbb{P}} 0.$$

Now, using the convergence results from [2], summarized in Proposition 2, we are able to determine the limiting random variables \mathcal{D}^α and \mathcal{D}^β .

PROPOSITION 4. *Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of CM graphs with $|V_n| = n$ and $\alpha, \beta \in \{+, -\}$. Then there exist integer valued random variables \mathcal{D}^α and \mathcal{D}^β such that for each $k, l \in \mathbb{N}$, as $n \rightarrow \infty$,*

$$\mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right] \xrightarrow{\mathbb{P}} \mathbb{P}(\mathcal{D}^\alpha = k) \mathbb{P}(\mathcal{D}^\beta = l).$$

PROOF. First let $(\alpha, \beta) = (+, -)$. Then it follows from Lemma 1 i) that

$$\begin{aligned} & \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right] = \\ & = \sum_{v, w \in V_n} I \{ \widehat{D}_n^+ v = k \} I \{ \widehat{D}_n^- w = l \} \frac{\widehat{D}_n^+ v \widehat{D}_n^- w}{|\widehat{E}_n|^2} = \\ & = \left(\sum_{v \in V_n} I \{ \widehat{D}_n^+ v = k \} \frac{\widehat{D}_n^+ v}{|\widehat{E}_n|} \right) \left(\sum_{w \in V_n} I \{ \widehat{D}_n^- w = l \} \frac{\widehat{D}_n^- w}{|\widehat{E}_n|} \right) = \\ & = \left(k \sum_{v \in V_n} \frac{I \{ \widehat{D}_n^+ v = k \}}{|\widehat{E}_n|} \right) \left(l \sum_{w \in V_n} \frac{I \{ \widehat{D}_n^- w = l \}}{|\widehat{E}_n|} \right) \xrightarrow{\mathbb{P}} \end{aligned}$$

$$\xrightarrow{\mathbb{P}} \frac{k\mathbb{P}(\xi = k)}{\mathbb{E}[\xi]} \frac{l\mathbb{P}(\gamma = l)}{\mathbb{E}[\gamma]} \quad \text{as } n \rightarrow \infty,$$

where the convergence in the last line is by Proposition 2. The other three cases are slightly more involved. Consider, for example, $(\alpha, \beta) = (-, +)$. Then we have,

$$\begin{aligned} \mathbb{E}\left[\mathbb{E}\left[I_{\xi_n^{\alpha,\beta}}(k, l) \mid G_n\right] \mid \widehat{\mathcal{D}}(G_n)\right] &= \tag{5.8} \\ &= \sum_{v \in V_n} I\left\{\widehat{D}_n^-(v) = k\right\} \frac{\widehat{D}_n^+(v)}{|\widehat{E}_n|} \sum_{w \in V_n} I\left\{\widehat{D}_n^+(w) = l\right\} \frac{\widehat{D}_n^-(w)}{|\widehat{E}_n|}. \end{aligned}$$

We will first analyze the last summation.

$$\begin{aligned} \frac{1}{|\widehat{E}_n|} \sum_{w \in V_n} \widehat{D}_n^-(w) I\left\{\widehat{D}_n^+(w) = l\right\} &= \frac{1}{|\widehat{E}_n|} \sum_{i \in \mathbb{N}} i \sum_{w \in V_n} I\left\{\widehat{D}_n^-(w) = i\right\} I\left\{\widehat{D}_n^+(w) = l\right\} \xrightarrow{\mathbb{P}} \\ &\xrightarrow{\mathbb{P}} \frac{\mathbb{P}(\xi = l)}{\mathbb{E}[\xi]} \sum_{i \in \mathbb{N}} i\mathbb{P}(\gamma = i) \quad \text{as } n \rightarrow \infty = \\ &= \frac{\mathbb{P}(\xi = l)\mathbb{E}[\gamma]}{\mathbb{E}[\xi]} = \mathbb{P}(\xi = l), \tag{5.9} \end{aligned}$$

where we again used Proposition 2 and $\mathbb{E}[\gamma] = \mathbb{E}[\xi]$. In a similar way we obtain that, as $n \rightarrow \infty$,

$$\frac{1}{|\widehat{E}_n|} \sum_{v \in V_n} \widehat{D}_n^+(v) I\left\{\widehat{D}_n^-(v) = k\right\} \xrightarrow{\mathbb{P}} \mathbb{P}(\gamma = k). \tag{5.10}$$

Applying (5.9) and (5.10) to (5.8) we get

$$\mathbb{E}\left[\mathbb{E}\left[I_{\xi_n^{-,+}}(k, l) \mid G_n\right] \mid \widehat{\mathcal{D}}(G_n)\right] \xrightarrow{\mathbb{P}} \mathbb{P}(\gamma = k)\mathbb{P}(\xi = l).$$

For the other two cases we obtain, as $n \rightarrow \infty$,

$$\begin{aligned} \mathbb{E}\left[\mathbb{E}\left[I_{\xi_n^{+,+}}(k, l) \mid G_n\right] \mid \widehat{\mathcal{D}}(G_n)\right] &\xrightarrow{\mathbb{P}} \frac{k\mathbb{P}(\xi = k)\mathbb{P}(\xi = l)}{\mathbb{E}[\xi]}, \\ \mathbb{E}\left[\mathbb{E}\left[I_{\xi_n^{-,-}}(k, l) \mid G_n\right] \mid \widehat{\mathcal{D}}(G_n)\right] &\xrightarrow{\mathbb{P}} \frac{l\mathbb{P}(\gamma = k)\mathbb{P}(\gamma = l)}{\mathbb{E}[\gamma]}. \end{aligned}$$

Table 1

Distributions of \mathcal{D}^α and \mathcal{D}^β for $\alpha, \beta \in \{+, -\}$.

α	β	$\mathbb{P}(\mathcal{D}^\alpha = k)$	$\mathbb{P}(\mathcal{D}^\beta = l)$
+	-	$k\mathbb{P}(\xi = k) / \mathbb{E}[\xi]$	$l\mathbb{P}(\gamma = l) / \mathbb{E}[\gamma]$
-	+	$\mathbb{P}(\gamma = k)$	$\mathbb{P}(\xi = l)$
+	+	$k\mathbb{P}(\xi = k) / \mathbb{E}[\xi]$	$\mathbb{P}(\xi = l)$
-	-	$\mathbb{P}(\gamma = k)$	$l\mathbb{P}(\gamma = l) / \mathbb{E}[\gamma]$.

The results now holds if we define \mathcal{D}^α and \mathcal{D}^β by their probabilities summarized in Table 1. □

We end this section with a convergence result for first and second moment of $\mathbb{E}\left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n\right]$.

PROPOSITION 5. *Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of CM graphs with $|V_n| = n$ and $\alpha, \beta \in \{+, -\}$. Then, for each $k, l \in \mathbb{N}$,*

- i) $\lim_{n \rightarrow \infty} \mathbb{E}\left[\mathbb{E}\left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n\right]\right] = \mathbb{P}(\mathcal{D}^\alpha = k) \mathbb{P}(\mathcal{D}^\beta = l),$
- ii) $\lim_{n \rightarrow \infty} \mathbb{E}\left[\mathbb{E}\left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n\right]^2\right] = \mathbb{P}(\mathcal{D}^\alpha = k)^2 \mathbb{P}(\mathcal{D}^\beta = l)^2,$

and hence, as $n \rightarrow \infty, \mathbb{E}\left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n\right] \xrightarrow{\mathbb{P}} \mathbb{P}(\mathcal{D}^\alpha = k) \mathbb{P}(\mathcal{D}^\beta = l).$

PROOF.

- i) Let $k, l \in \mathbb{N}$, then, since

$$\mathbb{E}\left[\mathbb{E}\left[I_{\mathcal{E}}^{\alpha,\beta}(k, l) \mid G_n\right] \mid \widehat{\mathcal{D}}(G_n)\right] \leq 1, \tag{5.11}$$

it follows, using Proposition 4 and dominated convergence, that for each pair $\alpha, \beta \in \{+, -\}$, we have

$$\lim_{n \rightarrow \infty} \mathbb{E}\left[\mathbb{E}\left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n\right]\right] = \mathbb{P}(\mathcal{D}^\alpha = k) \mathbb{P}(\mathcal{D}^\beta = l),$$

where $\mathcal{D}^\alpha, \mathcal{D}^\beta$ have distributions defined in Table 1.

ii) For the second moment we get, using conditioning on $\widehat{\mathcal{D}}(G_n)$,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right]^2 \right] &= \lim_{n \rightarrow \infty} \mathbb{E} \left[\mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right]^2 \mid \widehat{\mathcal{D}}(G_n) \right] \right] = \\ &= \lim_{n \rightarrow \infty} \mathbb{E} \left[\left(\sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \frac{\widehat{D}_n^+ \pi_* e \widehat{D}_n^- \pi^* e}{|\widehat{E}_n|^2} \right)^2 + o_{\mathbb{P}}(1) \right] = \end{aligned} \tag{5.12}$$

$$= \lim_{n \rightarrow \infty} \mathbb{E} \left[\mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}}^{\alpha, \beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right]^2 + o_{\mathbb{P}}(1) \right] = \tag{5.13}$$

$$= \left(\mathbb{P}(\mathcal{D}^\alpha = k) \mathbb{P}(\mathcal{D}^\beta = l) \right)^2. \tag{5.14}$$

Here (5.12) follows from Lemma 1 ii), (5.13) is by Lemma 1 i), and (5.14) is due to Proposition 4, continuous mapping theorem, (5.11) and the fact that the $o_{\mathbb{P}}(1)$ terms are uniformly bounded, see proof Lemma 1. The distributions of \mathcal{D}^α , \mathcal{D}^β are again given in Table 1.

The last result now follows by a second moment argument. □

5.2. Repeated Configuration Model

Described in Section 4.1 of [2], RCM connects inbound and outbound stubs uniformly at random and then the resulting graph is checked to be simple. If not, one repeats the connection step until the resulting graph is simple. If the distributions F_- and F_+ have finite variances, then the probability of the graph being simple converges to a non-zero number, see [2], Theorem 4.3. Therefore, throughout this section, we will assume that $\mathbb{E}[\gamma^2], \mathbb{E}[\xi^2] < \infty$.

Let $\{G_n\}_{n \in \mathbb{N}}$ be again a sequence of CM graphs, and let S_n denote the event that G_n is simple. We will prove, in Theorem 3 below, that for a sequence of RCM graphs of growing size, our three rank correlation measures converge to zero, by showing that for all $\alpha, \beta \in \{+, -\}$ and $k, l \in \mathbb{N}$,

$$\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n, S_n \right] \xrightarrow{\mathbb{P}} \mathbb{P}(\mathcal{D}^\alpha = k) \mathbb{P}(\mathcal{D}^\beta = l),$$

as $n \rightarrow \infty$, where \mathcal{D}^α and \mathcal{D}^β are random variables whose distributions are defined in Table 1.

First we show that, asymptotically, conditioning on the graph being simple does not effect the conditional expectation $\mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right]$.

LEMMA 2. Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of CM graphs with $|V_n| = n$ and $\alpha, \beta \in \{+, -\}$ and denote by S_n the event that G_n is simple. Then, for each $k, l \in \mathbb{N}$, as $n \rightarrow \infty$,

$$\left| \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n, S_n \right] \mid \widehat{\mathcal{D}}(G_n) \right] - \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right] \right| \xrightarrow{\mathbb{P}} 0.$$

PROOF. First, we write

$$\begin{aligned} & \left| \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n, S_n \right] \mid \widehat{\mathcal{D}}(G_n) \right] - \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right] \right| = \\ & = \left| \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right] \left(\frac{I \{S_n\}}{\mathbb{P}(S_n)} - 1 \right) \mid \widehat{\mathcal{D}}(G_n) \right] \right|. \end{aligned} \tag{5.15}$$

Next, denote by

$$\text{Var} \left(\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right) \quad \text{and} \quad \text{Var} \left(I \{S_n\} \mid \widehat{\mathcal{D}}(G_n) \right)$$

the variance of, respectively $\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right]$ and $I \{S_n\}$, conditioned on $\widehat{\mathcal{D}}(G_n)$. Then, by adding and subtracting in (5.15) the product of the conditional expectations

$$\mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right] \left(\frac{\mathbb{P} \left(S_n \mid \widehat{\mathcal{D}}(G_n) \right)}{\mathbb{P}(S_n)} - 1 \right),$$

we get

$$\begin{aligned} (5.15) & \leq \frac{1}{\mathbb{P}(S_n)} \sqrt{\text{Var} \left(I \{S_n\} \mid \widehat{\mathcal{D}}(G_n) \right)} \sqrt{\text{Var} \left(\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right)} + \\ & + \left| \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right] \left(\frac{\mathbb{P} \left(S_n \mid \widehat{\mathcal{D}}(G_n) \right)}{\mathbb{P}(S_n)} - 1 \right) \right| \leq \\ & \leq \frac{1}{\mathbb{P}(S_n)} \sqrt{\text{Var} \left(\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right)} + \left| \frac{\mathbb{P} \left(S_n \mid \widehat{\mathcal{D}}(G_n) \right)}{\mathbb{P}(S_n)} - 1 \right|. \end{aligned} \tag{5.16}$$

Following the argument in the first part of the proof of Proposition 4.4 from [2] we conclude that, $\mathbb{P} \left(S_n \mid \widehat{\mathcal{D}}(G_n) \right)$ and $\mathbb{P}(S_n)$ converge to the same positive limit,

hence the latter expression in (5.16) is $o_{\mathbb{P}}(1)$. The result now follows, since by Proposition 3

$$\text{Var} \left(\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right) = o_{\mathbb{P}}(1). \quad \square$$

In the next theorem we show that the conditions of Theorem 2 hold for a sequence of RCM graphs, and thus obtain the desired convergence of the three rank correlations, using a second moment argument.

THEOREM 3. *Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of RCM graphs with $|V_n| = n$ and $\alpha, \beta \in \{+, -\}$. Then, as $n \rightarrow \infty$,*

$$\rho_{\alpha}^{\beta}(G_n) \xrightarrow{\mathbb{P}} 0, \quad \bar{\rho}_{\alpha}^{\beta}(G_n) \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \tau_{\alpha}^{\beta}(G_n) \xrightarrow{\mathbb{P}} 0.$$

PROOF. Instead of conditioning on RCM graphs we condition on CM graphs G_n and the event that it is simple, S_n . Let $k, l \in \mathbb{N}$ and let $\mathcal{D}^{\alpha}, \mathcal{D}^{\beta}$ have distributions defined in Table 1. Then, for each pair $\alpha, \beta \in \{+, -\}$, we have

$$\begin{aligned} & \left| \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n, S_n \right] \mid \widehat{\mathcal{D}}(G_n) \right] - \mathbb{P}(\mathcal{D}^{\alpha} = k) \mathbb{P}(\mathcal{D}^{\beta} = l) \right| \leq \\ & \leq \left| \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n, S_n \right] \mid \widehat{\mathcal{D}}(G_n) \right] - \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right] \right| + \\ & + \left| \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right] - \mathbb{P}(\mathcal{D}^{\alpha} = k) \mathbb{P}(\mathcal{D}^{\beta} = l) \right|. \end{aligned}$$

Hence by Lemma 2 and Proposition 4 it follows that, as $n \rightarrow \infty$,

$$\mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n, S_n \right] \mid \widehat{\mathcal{D}}(G_n) \right] \xrightarrow{\mathbb{P}} \mathbb{P}(\mathcal{D}^{\alpha} = k) \mathbb{P}(\mathcal{D}^{\beta} = l).$$

Since $\mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n, S_n \right] \mid \widehat{\mathcal{D}}(G_n) \right] \leq 1$, dominated convergence and the above imply that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n, S_n \right] \right] = \mathbb{P}(\mathcal{D}^{\alpha} = k) \mathbb{P}(\mathcal{D}^{\beta} = l). \quad (5.17)$$

For the second moment we have

$$\left| \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n, S_n \right]^2 \mid \widehat{\mathcal{D}}(G_n) \right] - \mathbb{P}(\mathcal{D}^{\alpha} = k)^2 \mathbb{P}(\mathcal{D}^{\beta} = l)^2 \right| \leq$$

$$\leq \left| \mathbb{E} \left[\left(\left(\frac{I\{S_n\}}{\mathbb{P}(S_n)} \right)^2 - 1 \right) \mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n \right]^2 \mid \widehat{\mathcal{D}}(G_n) \right] \right| + \tag{5.18}$$

$$+ \left| \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n \right]^2 \mid \widehat{\mathcal{D}}(G_n) \right] - \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right]^2 \right| + \tag{5.19}$$

$$+ \left| \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n) \right]^2 - \mathbb{P}(\mathcal{D}^\alpha = k)^2 \mathbb{P}(\mathcal{D}^\beta = l)^2 \right|. \tag{5.20}$$

From Proposition 3 it follows that (5.19) converges to zero, while this holds for (5.20) because of Proposition 4 and the continuous mapping theorem. Finally, since

$$\left(\left(\frac{I\{S_n\}}{\mathbb{P}(S_n)} \right)^2 - 1 \right) \leq \left(\frac{I\{S_n\}}{\mathbb{P}(S_n)} - 1 \right) \left(1 + \mathbb{P}(S_n)^{-1} \right)$$

and $\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n \right] \leq 1,$

it follows that

$$(5.18) \leq \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n \right] \left(\frac{I\{S_n\}}{\mathbb{P}(S_n)} - 1 \right) \mid \widehat{\mathcal{D}}(G_n) \right] \left(1 + \mathbb{P}(S_n)^{-1} \right) \xrightarrow{\mathbb{P}} 0$$

as $n \rightarrow \infty,$

by (5.15), Lemma 2 and Proposition 4.4 from [2]. Therefore, using (5.11) and dominated convergence, we get

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n, S_n \right]^2 \right] = \mathbb{P}(\mathcal{D}^\alpha = k)^2 \mathbb{P}(\mathcal{D}^\beta = l)^2. \tag{5.21}$$

Combining (5.17) and (5.21), a second moment argument now yields that,

$$\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha,\beta}(k, l) \mid G_n, S_n \right] \xrightarrow{\mathbb{P}} \mathbb{P}(\mathcal{D}^\alpha = k) \mathbb{P}(\mathcal{D}^\beta = l) \quad \text{as } n \rightarrow \infty.$$

The result now follows from Theorem 2 by observing that the random variables \mathcal{D}^α and \mathcal{D}^β are independent and not concentrated in a single point. The latter is needed so that in case of average ranking we have $S_{\mathcal{D}^\alpha}(\mathcal{D}^\alpha) \neq 0,$ see Theorem 2. □

5.3. Erased Configuration Model

When the variances of the degree distributions are infinite, the probability of getting a simple graph using RCM converges to zero as the graph size increases. To remedy this we use ECM, described in Section 4.2 of [2]. In ECM stubs are connected at random, and then self-loops are removed and multiple edges are merged. We emphasize that for this model the actual degree sequence $\mathfrak{D}(G)$ may differ from the bi-degree sequence, $\widehat{\mathfrak{D}}(G)$, used to do the pairing.

We will often use results from Proposition 4.5 of [2], which we state below for reference.

PROPOSITION 6 ([2], PROPOSITION 4.5). *Let $G_n = (V_n, E_n)$ be a sequence of ECM graphs with $|V_n| = n$ and $k, l \in \mathbb{N}$. Then, as $n \rightarrow \infty$,*

$$\frac{1}{n} \sum_{v \in V_n} I \{D^+ v = k\} \xrightarrow{\mathbb{P}} \mathbb{P}(\xi = k) \quad \text{and} \quad \frac{1}{n} \sum_{v \in V_n} I \{D^- v = l\} \xrightarrow{\mathbb{P}} \mathbb{P}(\gamma = l).$$

We will follow the same second moment argument approach as in the previous section to prove that all three rank correlations, ρ , $\bar{\rho}$ and τ converge to zero in ECM. First we will establish a convergence result for the total number of erased in- and outbound stubs.

For $v, w \in V$ and $\alpha \in \{+, -\}$, we denote by $E^{c, \alpha}(v)$ and $E^c(v, w)$, respectively, the set of erased α -stubs from v and erased edges between v and w . For $e \in V^2$, we write $E^c(e) = E^c(\pi_* e, \pi^* e)$.

LEMMA 3. *Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of ECM graphs with $|V_n| = n$ and $\alpha \in \{+, -\}$. Then*

$$\frac{1}{n} \sum_{v \in V_n} |E_n^{c, \alpha}(v)| \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \rightarrow \infty.$$

PROOF. Let $N \in \mathbb{N}$ and fix a $v \in V_N$, then for all $n \geq N$, $|E_n^{c, \alpha}(v)| \leq \gamma_n + 1$ where all γ_n are i. i. d. copies of γ . Since by Lemma 5.2 from [2] we have $E_n^{c, \alpha}(v) \rightarrow 0$ almost surely and furthermore $\mathbb{E}[\gamma] < \infty$, dominated convergence implies that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{v \in V_n} \mathbb{E}[|E_n^{c, \alpha}(v)|] = 0.$$

Applying the Markov inequality then yields, for arbitrary $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{n} \sum_{v \in V_n} |E_n^{c, \alpha}(v)| \geq \varepsilon \right) \leq \lim_{n \rightarrow \infty} \frac{\sum_{v \in V_n} \mathbb{E}[|E_n^{c, \alpha}(v)|]}{n\varepsilon} = 0. \quad \square$$

Since

$$|E| = |\widehat{E}| - \sum_{v \in V} |E^{c, \alpha}(v)| \quad \text{for } \alpha \in \{+, -\},$$

the above lemma combined with Proposition 2 implies that

$$\frac{|E_n|}{n} \xrightarrow{\mathbb{P}} \mathbb{E}[\gamma] \quad \text{as } n \rightarrow \infty. \tag{5.22}$$

We proceed with the next lemma, which is an adjustment of Lemma 1, where we now condition on both the bi-degree sequence of stubs as well as the eventual degree sequence. We remark that $I_e^{\alpha, \beta}(k, l)$ is completely determined by the latter while $\sum_{e \in V^2} |E^c(e)|$ is completely determined by the combination of the two sequences. Recall that for $e \in V^2$, $|\widehat{E}(e)|$ denotes the number of edges $f \in E$ with $f = e$ before removal of self-loops and merging multiple edges and observe that $|E(e)| = |\widehat{E}(e)| - |E^c(e)|$.

LEMMA 4. *Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of ECM graphs with $|V_n| = n$. Then, for each $k, l \in \mathbb{N}$ and $\alpha, \beta \in \{+, -\}$,*

$$\begin{aligned} \text{i) } & \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n), \mathcal{D}(G_n) \right] = \\ & = \sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \frac{D_n^+ \pi_* e D_n^- \pi^* e}{|E_n|^2} + o_{\mathbb{P}}(1), \\ \text{ii) } & \mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right]^2 \mid \widehat{\mathcal{D}}(G_n), \mathcal{D}(G_n) \right] = \\ & = \left(\sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \frac{D_n^+ \pi_* e D_n^- \pi^* e}{|E_n|^2} \right)^2 + o_{\mathbb{P}}(1). \end{aligned}$$

To obtain this result we need the following Lemma.

LEMMA 5. Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of ECM graphs with $|V_n| = n$. Then, for each $k, l \in \mathbb{N}$ and $\alpha, \beta \in \{+, -\}$,

$$\sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \frac{\widehat{D}_n^+ \pi_* e \widehat{D}_n^- \pi_* e}{|\widehat{E}_n|^2} = \sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \frac{D_n^+ \pi_* e D_n^- \pi_* e}{|\widehat{E}_n|^2} + o_{\mathbb{P}}(1).$$

PROOF. Since $\widehat{D}_n^\alpha \pi e = D_n^\alpha \pi e + |E_n^{c, \alpha}(\pi e)|$, we have

$$\begin{aligned} \sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \frac{\widehat{D}_n^+ \pi_* e \widehat{D}_n^- \pi_* e}{|\widehat{E}_n|^2} &= \sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \frac{D_n^+ \pi_* e D_n^- \pi_* e}{|\widehat{E}_n|^2} + \\ &+ \sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \frac{\widehat{D}_n^+ \pi_* e |E_n^{c, -}(\pi_* e)|}{|\widehat{E}_n|^2} + \end{aligned} \quad (5.23)$$

$$+ \sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \frac{\widehat{D}_n^- \pi_* e |E_n^{c, +}(\pi_* e)|}{|\widehat{E}_n|^2} + \quad (5.24)$$

$$+ \sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \frac{|E_n^{c, +}(\pi_* e)| |E_n^{c, -}(\pi_* e)|}{|\widehat{E}_n|^2}. \quad (5.25)$$

By Lemma 3 and Proposition 2 it follows that (5.25) is $o_{\mathbb{P}}(1)$. For (5.23) we have

$$\begin{aligned} \sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \frac{\widehat{D}_n^+ \pi_* e |E_n^{c, -}(\pi_* e)|}{|\widehat{E}_n|^2} &\leq \sum_{v \in V_n} \frac{\widehat{D}_n^+ v}{|\widehat{E}_n|} \sum_{w \in V_n} \frac{|E_n^{c, -}(w)|}{|\widehat{E}_n|} \leq \\ &\leq \sum_{w \in V_n} \frac{|E_n^{c, -}(w)|}{|\widehat{E}_n|} = o_{\mathbb{P}}(1), \end{aligned}$$

where the last line is due to $\sum_{v \in V_n} \widehat{D}_n^+ v = |\widehat{E}_n|$. The last equation then follows from Lemma 3 and Proposition 2. This holds similarly for (5.24) and hence the result follows. \square

PROOF OF LEMMA 4.

i) By splitting $|E_n(e)|$ we obtain,

$$\mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n), \mathcal{D}(G_n) \right] =$$

$$\begin{aligned}
&= \mathbb{E} \left[\sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \frac{|E_n(e)|}{|E_n|} \middle| \widehat{\mathcal{D}}(G_n), \mathcal{D}(G_n) \right] = \\
&= \frac{|\widehat{E}_n|}{|E_n|} \mathbb{E} \left[\sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \frac{|\widehat{E}_n(e)|}{|\widehat{E}_n|} \middle| \widehat{\mathcal{D}}(G_n) \right] + \tag{5.26}
\end{aligned}$$

$$- \frac{1}{|E_n|} \sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \mathbb{E} \left[|E_n^c(e)| \middle| \widehat{\mathcal{D}}(G_n), \mathcal{D}(G_n) \right]. \tag{5.27}$$

For (5.27) we have,

$$\begin{aligned}
\frac{1}{|E_n|} \sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \mathbb{E} \left[|E_n^c(e)| \middle| \widehat{\mathcal{D}}(G_n), \mathcal{D}(G_n) \right] &\leq \\
&\leq \frac{1}{|E_n|} \sum_{e \in V_n^2} \mathbb{E} \left[|E_n^c(e)| \middle| \widehat{\mathcal{D}}(G_n), \mathcal{D}(G_n) \right] = \\
&= \frac{1}{|E_n|} \sum_{v \in V_n} |E_n^{c,+}(v)|,
\end{aligned}$$

which is $o_{\mathbb{P}}(1)$ by Lemma 3 and (5.22). Now, since the conditional expectation in (5.26) equals (5.1), it follows from Lemma 1 i), Lemma 5 and (5.22) that

$$(5.26) = \sum_{e \in V_n^2} I_e^{\alpha, \beta}(k, l) \frac{D_n^+ \pi_* e D_n^- \pi^* e}{|E_n|^2} + o_{\mathbb{P}}(1).$$

ii) Splitting both terms $|E_n(e)|$ and $|E_n(f)|$ for $e, f \in V_n^2$ yields,

$$\begin{aligned}
&\mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \middle| G_n \right]^2 \middle| \widehat{\mathcal{D}}(G_n), \mathcal{D}(G_n) \right] = \\
&= \mathbb{E} \left[\sum_{e, f \in V_n^2} I_e^{\alpha, \beta}(k, l) I_f^{\alpha, \beta}(k, l) \frac{|E_n(e)| |E_n(f)|}{|E_n|^2} \middle| \widehat{\mathcal{D}}(G_n), \mathcal{D}(G_n) \right] = \\
&= \frac{|\widehat{E}_n|^2}{|E_n|^2} \mathbb{E} \left[\sum_{e, f \in V_n^2} I_e^{\alpha, \beta}(k, l) I_f^{\alpha, \beta}(k, l) \frac{|\widehat{E}_n(e)| |\widehat{E}_n(f)|}{|\widehat{E}_n|} \middle| \widehat{\mathcal{D}}(G_n) \right] + \tag{5.28}
\end{aligned}$$

$$+ \sum_{e,f \in V_n^2} I_e^{\alpha,\beta}(k,l) I_f^{\alpha,\beta}(k,l) \mathbb{E} \left[\frac{|E_n^c(e)||E_n^c(f)|}{|E_n|^2} \middle| \widehat{\mathfrak{D}}(G_n), \mathfrak{D}(G_n) \right] + \quad (5.29)$$

$$- \sum_{e,f \in V_n^2} I_e^{\alpha,\beta}(k,l) I_f^{\alpha,\beta}(k,l) \mathbb{E} \left[\frac{|E_n^c(e)||\widehat{E}_n(f)|}{|E_n|^2} \middle| \widehat{\mathfrak{D}}(G_n), \mathfrak{D}(G_n) \right] + \quad (5.30)$$

$$- \sum_{e,f \in V_n^2} I_e^{\alpha,\beta}(k,l) I_f^{\alpha,\beta}(k,l) \mathbb{E} \left[\frac{|E_n^c(f)||\widehat{E}_n(e)|}{|E_n|^2} \middle| \widehat{\mathfrak{D}}(G_n), \mathfrak{D}(G_n) \right]. \quad (5.31)$$

Recognizing the conditional expectation in (5.28) as (5.2), then using first Lemma 1 ii) and then Lemma 5 and (5.22), it follows that (5.28) equals

$$\left(\sum_{e \in V_n^2} I_e^{\alpha,\beta}(k,l) \frac{D_n^+ \pi_* e D_n^- \pi^* e}{|E_n|^2} \right)^2 + o_{\mathbb{P}}(1).$$

It remains to show that (5.29)-(5.31) are $o_{\mathbb{P}}(1)$. For (5.29) we have

$$\begin{aligned} \sum_{e,f \in V_n^2} I_e^{\alpha,\beta}(k,l) I_f^{\alpha,\beta}(k,l) \mathbb{E} \left[\frac{|E_n^c(e)||E_n^c(f)|}{|E_n|^2} \middle| \widehat{\mathfrak{D}}(G_n), \mathfrak{D}(G_n) \right] &\leq \\ &\leq \left(\frac{1}{|E_n|} \sum_{v \in V_n} |E_n^{c,+}(v)| \right)^2 = o_{\mathbb{P}}(1) \end{aligned}$$

by Lemma 3 and (5.22). Since (5.30) and (5.31) are symmetric we will only consider the latter:

$$\begin{aligned} \sum_{e,f \in V_n^2} I_e^{\alpha,\beta}(k,l) I_f^{\alpha,\beta}(k,l) \mathbb{E} \left[\frac{|E_n^c(f)||\widehat{E}_n(e)|}{|E_n|^2} \middle| \widehat{\mathfrak{D}}(G_n), \mathfrak{D}(G_n) \right] &\leq \\ &\leq \left(\sum_{f \in V_n^2} \frac{|E_n^c(f)|}{|E_n|} \right) \frac{1}{|E_n|} \sum_{e \in V_n^2} \mathbb{E} \left[|\widehat{E}_n(e)| \middle| \widehat{\mathfrak{D}}(G_n) \right] = \\ &= \left(\sum_{v \in V_n} \frac{|E_n^+(v)|}{|E_n|} \right) \frac{|\widehat{E}_n|}{|E_n|} = o_{\mathbb{P}}(1). \end{aligned}$$

Here, for the last line, we used $\sum_{e \in V_n^2} \mathbb{E} \left[|\widehat{E}_n(e)| \middle| \widehat{\mathfrak{D}}(G_n) \right] = |\widehat{E}_n|$, and then Lemma 3 and (5.22). □

A straightforward adaptation of the proof of Proposition 4, using Lemma 4 instead of Lemma 1, yields the following result.

PROPOSITION 7. *Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of ECM graphs with $|V_n| = n$ and $\alpha, \beta \in \{+, -\}$. Then there exist integer valued random variables \mathcal{D}^α and \mathcal{D}^β such that for each $k, l \in \mathbb{N}$, as $n \rightarrow \infty$,*

$$\mathbb{E} \left[\mathbb{E} \left[I_{\mathcal{E}_n}^{\alpha, \beta}(k, l) \mid G_n \right] \mid \widehat{\mathcal{D}}(G_n), \mathcal{D}(G_n) \right] \xrightarrow{\mathbb{P}} \mathbb{P}(\mathcal{D}^\alpha = k) \mathbb{P}(\mathcal{D}^\beta = l),$$

where the distributions of \mathcal{D}^α and \mathcal{D}^β are given in Table 1.

We can now again use a second moment argument to get the convergence result for the three rank correlations in the Erased Configuration Model. We omit the proof since the computation of the variance follows the exact same steps as those in Proposition 5, where now, instead of only conditioning on $\widehat{\mathcal{D}}(G_n)$, we also condition on $\mathcal{D}(G_n)$ and use Lemma 4.

THEOREM 4. *Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of ECM graphs with $|V_n| = n$ and $\alpha, \beta \in \{+, -\}$. Then, as $n \rightarrow \infty$,*

$$\rho_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} 0, \quad \bar{\rho}_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \tau_\alpha^\beta(G_n) \xrightarrow{\mathbb{P}} 0.$$

This theorem shows that even when the variance of the degree sequences is infinite, one can construct a random graph for which the degree-degree dependencies, measured by rank correlations, converge to zero in the infinite graph size limit. Therefore this model can be used as a null model for such dependencies.

Acknowledgments

We like to thank an anonymous referee for thoroughly reading our manuscript and giving constructive comments and suggestions for improvement.

This work is supported by the EU-FET Open grant NADINE (288956).

6. Appendix A Continuization

In this appendix we will establish several relations between the distribution functions of integer valued random variables and their continuizations, using the functions \mathcal{F} and \mathcal{H} defined in (2.1) and (2.2), respectively.

Let $\tilde{X} = X + U$ be as in Definition 2, take $k \in \mathbb{Z}$ and define $I_k = [k, k + 1)$. Then for $x \in I_k$,

$$F_{\tilde{X}}(x) = (x - k)F_X(k) + (k + 1 - x)F_X(k - 1). \quad (6.1)$$

As a consequence, it follows that for $x \in I_k$,

$$dF_{\tilde{X}}(x) = (F_X(k) - F_X(k - 1)) dx = \mathbb{P}(X = k) dx. \quad (6.2)$$

These identities capture the essential relations between X and its continuization \tilde{X} . As a first result we have the following.

LEMMA 6. *Let X be an integer valued random variable and $m \in \mathbb{N}$. Then,*

$$\mathbb{E}\left[F_{\tilde{X}}(\tilde{X})^m\right] = \frac{1}{m + 1} \sum_{i=0}^m \mathbb{E}\left[F_X(X)^i F_X(X - 1)^{m-i}\right].$$

PROOF. Using (6.1) we obtain,

$$\begin{aligned} \int_{I_k} F_{\tilde{X}}(x)^m dx &= \int_{I_k} ((x - k)F_X(k) + (k + 1 - x)F_X(k - 1))^m dx = \\ &= \sum_{i=0}^m \binom{m}{i} F_X(k)^i F_X(k - 1)^{m-i} \int_0^1 (y)^i (1 - y)^{m-i} dy = \\ &= \sum_{i=0}^m \frac{m!}{i!(m - i)!} F_X(k)^i F_X(k - 1)^{m-i} \frac{\Gamma(i + 1)\Gamma(m - i + 1)}{\Gamma(m + 2)} = \\ &= \frac{1}{m + 1} \sum_{i=0}^m F_X(k)^i F_X(k - 1)^{m-i}, \end{aligned}$$

Combining this with (6.2), we get

$$\begin{aligned} \mathbb{E}\left[F_{\tilde{X}}(\tilde{X})^m\right] &= \sum_{k \in \mathbb{Z}} \int_{I_k} F_{\tilde{X}}(x)^m dF_{\tilde{X}}(x) = \\ &= \sum_{k \in \mathbb{Z}} \int_{I_k} F_{\tilde{X}}(x)^m \mathbb{P}(X = k) dx = \end{aligned}$$

$$= \frac{1}{m+1} \sum_{i=0}^m \mathbb{E} \left[F_X(X)^i F_X(X-1)^{m-i} \right]. \quad \square$$

As a direct consequence of Lemma 6 we get

$$\frac{1}{2} = \mathbb{E} \left[F_{\tilde{X}}(\tilde{X}) \right] = \frac{1}{2} \mathbb{E} [\mathcal{F}_X(X)], \quad (6.3)$$

relating $F_{\tilde{X}}$ to \mathcal{F}_X . Similar to (6.1), if Z is a random element independent of X , we get for $x \in I_k$,

$$F_{\tilde{X}|Z}(x) = (x-k)F_{X|Z}(k) + (k+1-x)F_{X|Z}(k-1). \quad (6.4)$$

Applying (6.4) in a similar way as (6.1) we arrive at an extension of Lemma 6. The proof is elementary, hence omitted.

PROPOSITION 8. *Let X be an integer valued random variable and Z a random element independent of the continuous part of \tilde{X} . Then*

- i) $\mathbb{E} \left[F_{\tilde{X}}(\tilde{X}) \mid Z \right] = \frac{1}{2} \mathbb{E} [\mathcal{F}_X(X) \mid Z],$ a.s.;
- ii) $F_{\tilde{X}|Z}(\tilde{X}) = \frac{1}{2} \mathcal{F}_{X|Z}(X),$ a.s.

The following results are extensions of the previous ones to the case of two integer valued random variables X and Y . We will state these without proofs, since these are either straightforward extensions of those for the case of a single random variable or follow from elementary calculations and the previous results.

LEMMA 7. *Let X, Y be integer valued random variables. Then,*

- i) $\mathbb{E} \left[F_{\tilde{X}}(\tilde{X}) F_{\tilde{Y}}(\tilde{Y}) \right] = \frac{1}{4} \mathbb{E} [\mathcal{F}_X(X) \mathcal{F}_Y(Y)],$
- ii) $\mathbb{E} \left[H_{\tilde{X}, \tilde{Y}}(\tilde{X}, \tilde{Y}) \right] = \frac{1}{4} \mathbb{E} [\mathcal{H}_{X,Y}(X, Y)].$

PROPOSITION 9. *Let X, Y be integer valued random variables and let Z be a random variable independent of the uniform parts of \tilde{X} and \tilde{Y} . Then*

- i) $\mathbb{E} \left[\tilde{F}_{\tilde{X}}(\tilde{X}) \tilde{F}_{\tilde{Y}}(\tilde{Y}) \mid Z \right] = \frac{1}{4} \mathbb{E} [\mathcal{F}_X(X) \mathcal{F}_Y(Y) \mid Z]$ a.s.;
- ii) $H_{\tilde{X}, \tilde{Y}|Z}(\tilde{X}, \tilde{Y}) = \frac{1}{4} \mathcal{H}_{X,Y|Z}(X, Y)$ a.s.

Bibliography

1. **M. Boguná, R. Pastor-Satorras, A. Vespignani**, *Epidemic spreading in complex networks with degree correlations*. arXiv preprint cond-mat/0301149, 2003.
2. **N. Chen, M. Olvera-Cravioto**, *Directed random graphs with given degree distributions*. Stochastic Systems, 3(1):147–186, 2013.
3. **F. Chung, L. Lu, V. Vu**, *Spectra of random graphs with given expected degrees*. Proceedings of the National Academy of Sciences, 100(11):6313–6318, 2003.
4. **M. Denuit, P. Lambert**, *Constraints on concordance measures in bivariate discrete data*. Journal of Multivariate Analysis, 93(1):40–57, 2005.
5. **S. N. Dorogovtsev, A. L. Ferreira, A. V. Goltsev, J. F. F. Mendes**, *Zero pearson coefficient for strongly correlated growing trees*. Physical Review E, 81(3):031135, 2010.
6. **T. Hasegawa, T. Takaguchi, N. Masuda**, *Observability transitions in correlated networks*. Physical Review E, 88(4):042809, 2013.
7. **N. Litvak, R. van der Hofstad** *Degree-degree correlations in random graphs with heavy-tailed degrees*. arXiv preprint arXiv:1202.3071, 2012.
8. **N. Litvak, R. van der Hofstad**, *Uncovering disassortativity in large scale-free networks*. Physical Review E, 87(2):022801, 2013.
9. **X. F. Liu, C. K. Tse**, *Impact of degree mixing pattern on consensus formation in social networks*. Physica A: Statistical Mechanics and its Applications, 407:1–6, 2014.
10. **S. Maslov, K. Sneppen, A. Zaliznyak**, *Detection of topological patterns in complex networks: correlation profile of the internet*. Physica A: Statistical Mechanics and its Applications, 333:529–540, 2004.
11. **M. Mesfioui, A. Tajar**, *On the properties of some nonparametric concordance measures in the discrete case*. Nonparametric Statistics, 17(5):541–554, 2005.
12. **M. E. J. Newman**, *Assortative mixing in networks*. Physical review letters, 89(20):208701, 2002.
13. **M. E. J. Newman**, *Mixing patterns in networks*. Physical Review E, 67(2):026126, 2003.
14. **J. Park, M. E. J. Newman**, *Origin of degree correlations in the internet and other networks*. Physical Review E, 68(2):026112, 2003.
15. **A. Srivastava, B. Mitra, F. Peruani, N. Ganguly**, *Attacks on correlated peer-to-peer networks: An analytical study*. pages 1076–1081, 2011.
16. **R. van der Hofstad**, *Random graphs and complex networks*. Unpublished manuscript, 2007.
17. **P. van der Hoorn, N. Litvak**, *Degree-degree correlations in directed networks with heavy-tailed degrees*. arXiv preprint arXiv:1310.6528, 2013.

PIM VAN DER HOORN

University of Twente
Faculty of Electrical Engineering, Mathematics
and Computer Sciences
Dep. Stochastic Operational Research
P.O. Box 217
7500 AE Enschede
The Netherlands
w.l.f.vanderhoorn@utwente.nl

NELLY LITVAK

University of Twente
Faculty of Electrical Engineering, Mathematics
and Computer Sciences
Dep. Stochastic Operational Research
P.O. Box 217
7500 AE Enschede
The Netherlands
N.Litvak@utwente.nl, n.litvak@utwente.nl