Asymptotic Behavior of Common Connections in Sparse Random Networks

Bikramjit Das¹ · Tiandong Wang² · Gengling Dai¹

Received: 16 June 2021 / Revised: 31 August 2021 / Accepted: 18 September 2021 / Published online: 26 October 2021 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract
Random network models generated using sparse exchangeable graphs have provided a mechanism to study a wide variety of complex real-life networks. In particular, these models help with investigating power-law properties of degree distributions, number of edges, and other relevant network metrics which support the scale-free structure of networks. Previous work on such graphs imposes a marginal assumption of univariate regular variation (e.g., power-law tail) on the bivariate generating graphex function. In this paper, we study sparse exchangeable graphs generated by graphex functions which are multivariate regularly varying. We also focus on a different metric for our study: the distribution of the number of common vertices (connections) shared by a pair of vertices. The number being high for a fixed pair is an indicator of the original pair of vertices being connected. We find that the distribution of number of common connections are regularly varying as well, where the tail indices of regular variation are governed by the type of graphex function used. Our results are verified on simulated graphs by estimating the relevant tail index parameters.

Keywords Random networks · Common connections · Power laws · Multivariate regular variation

Mathematics Subject Classification 05C82 · 60F15 · 60G70

Bikramjit Das and Tiandong Wang contributed equally to this work.

* Tiandong Wang
twang@stat.tamu.edu
Bikramjit Das
bikram@sutd.edu.sg
Gengling Dai
gengling_dai@mymail.sutd.edu.sg

¹ Engineering Systems and Design, Singapore University of Technology and Design, 8 Somapah Road, 487372, Singapore
² Department of Statistics, Texas A&M University, 3143 TAMU, 77840 College Station, TX, US
1 Introduction

Modeling social, economic, and biological networks has become a principal area of interest for data scientists in the recent decades. The degree distribution provides an idea about the structure of the network, e.g., a power-law behavior here indicating the so-called “scale-free” structure; see Voitalov et al. (2019) for a recent discussion. Another quantity of potential interest to network scientists, especially in the context of social networking platforms like Facebook, Instagram, LinkedIn, and Twitter, is the number of common connections or common friends between two vertices. This can be thought of as a generalization of the degree of a vertex. Network recommendation systems use this number as one important metric to suggest potential new vertices (friends) to platform users (Gupta et al. 2013). The connection between link prediction and the size of common connections has been explored in the literature; see Liben-Nowell and Kleinberg (2007); Newman (2001); Barabási et al. (2002) for further discussions.

Although of interest to network scientists, only few theoretical results exist in the literature characterizing the asymptotic distribution of common connections for growing network models. While studying clustering coefficients for random intersection graphs, Bloznelis and Kurauskas (2016) find the asymptotic distribution of common connections for a randomly selected pair of vertices. In addition, for a certain class of linear preferential attachment models, Das and Ghosh (2021) identify the growth rate for common connections between two fixed vertices in the graph. In our paper, we study common connections in the context of the erstwhile popular and flexible framework of exchangeable random graph models called graphex processes.

Graphex models, or vertex-exchangeable random graphs have been used extensively to study statistical network models that are dense (Hoover (1979); Aldous (1981); Lovász and Szegedy (2006); Diaconis and Janson (2008)). Recent work has successfully extended the classical graphex framework to sparse networks, which mimics many real-world networks; see Caron and Fox (2017); Borgs et al. (2019). In particular, they point out that exchangeable random measures admit a representation theorem due to Kallenberg (1990), giving a general construction for such graph models. Moreover, Veitch and Roy (2015) and Borgs et al. (2018) show how such construction naturally generalizes the dense exchangeable graphex framework to the sparse regime, and analyze some of the properties of the associated class of random graphs, called graphex processes; also see for instance, Janson (2016, 2017), Veitch and Roy (2019) and Borgs et al. (2019). Furthermore, Herlau et al. (2016) and Todeschini et al. (2020) have developed sparse graph models with (overlapping) community structure within the graphex framework. However, none of the preceding work has focused on the asymptotic behavior of common connections in a sparse graphex model, which forms the main motivation of this paper.

In the sequel, we work under the framework of undirected, sparse, and vertex-exchangeable random graphs generated by graphex functions which are multivariate regularly varying, a modest generalization of power-law tailed functions, with formal definitions given in Sect. 2.2. The regular variation of the graphex function creates a sparse graph, and we show that the distribution of common connections in the generated graphs is also regularly varying, i.e., approximately power-law tailed. Moreover, we observe that depending on the type of the graphex function, i.e., whether it is a product of univariate functions (called separable), or not (called non-separable), we get different asymptotic behaviors of common connections in the random graph. We estimate necessary parameters in simulated random graphs to numerically verify and support the results obtained.
The remaining of the paper proceeds as follows. In Sect. 2, we introduce the random graph framework, as well as necessary definitions and assumptions of multivariate regularly varying functions in this context. The effects of different regular variation assumptions on particular marginals (i.e., univariate, bivariate, etc.) of the graphex functions are also characterized in this section. The main results on the asymptotic distribution of common connections are collected in Sect. 3. In addition, Sect. 4 presents results from simulation studies, where we provide typical examples of sparse random networks and verify our results of Sect. 3. Finally, concluding remarks are given in Sect. 5.

2 Random Graphs, Regular Variation and Graphex Functions

In this section, we create our framework for network models via graphex processes, discuss multivariate regular variation, and derive regular variation properties for functionals of graphex processes.

2.1 Random Graph Models

We now present a framework which models sparse random graphs using (infinite) Poisson point processes on \( \mathbb{R}^2_+ = [0, \infty)^2 \). Finite graphs are obtained by appropriately truncating the support of the point process. The model representation is based on the work of Caron and Fox (2017); Veitch and Roy (2015); Borgs et al. (2018); Caron et al. (2020). Here we follow their construction and notations. Such representation is also related to Kallenberg exchangeable graphs (Kallenberg 1990), and can be characterized by a symmetric measurable function, \( W : [0, \infty)^2 \mapsto [0, 1] \), often called a graphex function, which was originally introduced in Veitch and Roy (2015). A variety of sparse random graphs can be parametrized using the graphex function \( W \); we will ignore isolated vertices in the graph as is customarily done, see Caron and Fox (2017); Caron et al. (2020) for details.

Let \( \Pi = (\theta_i, \eta_i)_{\mathbb{N}} \) be a unit-rate Poisson process on \( \mathbb{R}_+^2 \). In this representation, vertices are embedded at some location \( \theta_i \in \mathbb{R}_+^2 \), and there is a latent variable \( \eta_i \in \mathbb{R}_+^2 \) which along with the graphex function \( W \) determines the edges of the graph. For this paper, we follow the definition of a graphex function as in Caron et al. (2020), which is outlined below.

**Definition 1** A symmetric measurable function \( W : [0, \infty)^2 \mapsto [0, 1] \) is a graphex function if it satisfies the following criteria:

(a) \( 0 < \overline{W} = \int_0^\infty \int_0^\infty W(x, y) \, dx \, dy < \infty \);
(b) \( \int_0^\infty W(x, x) \, dx < \infty \);
(c) \( \lim_{x \to 0} W(x, x) \) and \( \lim_{x \to \infty} W(x, x) \) both exist.

Since \( \overline{W} < \infty \), we have \( \lim_{x \to 0} W(x, x) = 0 \). The three criteria given in Definition 1 are chosen for technical reasons so that the function \( W \) is well-behaved, and we will impose other necessary regularity conditions on \( W \) later in the paper. Edges of the graph induced by \( \Pi \) and \( W \) are given by a point process:

\[
Z = \sum_{i,j} Z_{ij} \delta_{(\theta_i, \theta_j)}
\]
where \( Z_{ij} = Z_{ji} \) is a binary variable which takes the value 1 if there is an edge between \( \theta_i \) and \( \theta_j \) and 0 otherwise. Given the graphex function \( W \) and \( i \leq j \), we have
\[
Z_{ij} \Pi, W \sim \text{Bernoulli} \left( W(\eta_i, \eta_j) \right)
\]
and \{\( Z_{ij} : i \leq j \)\} are independent random variables. We obtain a finite size random graph family \( (G_t)_{t \geq 0} \), often called a graphex process, by truncating the process \( \sum_{i,j} Z_{ij} \delta_{\theta_i, \theta_j} \) to the square \([0, t]^2\). Specifically, let \( G_t = (\mathcal{E}_t, \mathcal{V}_t) \) denote a graph of size \( t \geq 0 \) with vertex set \( \mathcal{V}_t \) and edge set \( \mathcal{E}_t \), respectively given by
\[
\mathcal{V}_t = \{ \theta | \theta \leq t \text{ and } Z_{ik} = 1 \text{ for some } k \text{ with } \theta_k \leq t \}, \\
\mathcal{E}_t = \{ (\theta_i, \theta_j) | \theta_i, \theta_j \leq t \text{ and } Z_{ij} = 1 \}.
\]
The following quantities are of interest for studying properties of graphex processes: the univariate and bivariate graphex marginal functions \( \mu_1 : \mathbb{R}_+ \to \mathbb{R}_+ \) and \( \mu_2 : \mathbb{R}_+^2 \to \mathbb{R}_+ \) are given by
\[
\mu_1(x) := \int_0^\infty W(x, y) \, dy, 
\]
\[
\mu_2(x, y) := \int_0^\infty W(x, z) W(y, z) \, dz,
\]
respectively. The integrability of \( W \) in \( \mathbb{R}_+^2 \) guarantees that both \( \mu_1 \) and \( \mu_2 \) are well-defined and finite. We can check that \( \mu_1(\eta_i) \) is proportional to the average number of vertices that a vertex at \( \theta_i \) is connected to, whereas \( \mu_2(\eta_i, \eta_j) \) is proportional to the average number of common connections (vertices) between the two vertices at \( \theta_i \) and \( \theta_j \). We also further define the \( d \)-variate graphex marginal for \( d \geq 3 \) as:
\[
\mu_d(x_1, \ldots, x_d) := \int_0^\infty \prod_{i=1}^d W(x_i, z) \, dz, \quad x_i \neq x_j, \quad i \neq j, \quad i, j = 1, \ldots, d.
\]

### 2.2 Regular Variation

Studies involving the Kallenberg exchangeable sparse graphs have been occasionally carried out under the assumption that the univariate graphex marginal \( \mu_1(x) \) has a power-law-like tail (regularly varying) as \( x \to \infty \); see Naulet et al. (2021); Caron et al. (2020). Instead of imposing a condition on \( \mu_1 \), we start by imposing important conditions on the generating graphex function, \( W \), which is the direct input for the generation of the underlying network. It turns out that the regularity conditions on \( W \) are also closely connected to the theory of multivariate regular variation (Bingham et al. 1989; Resnick 2007), which we now outline.

A function \( f : (0, \infty) \to (0, \infty) \) is regularly varying (at infinity) if \( \lim_{t \to \infty} f(tx)/f(t) = x^\beta \) for \( x > 0 \) and \( \beta \in \mathbb{R} \). Here \( \beta \) is the tail index or index of regular variation and we write \( f \in \mathcal{R}_\beta \). In dimensions \( d > 1 \) this can be extended to multivariate regular variation (cf. Stam (1977); de Haan and Resnick (1987)), which we now present.

**Definition 2** Suppose \( \mathbb{C} \subset \mathbb{R}_+^d \) is a cone, that is, \( x \in \mathbb{C} \) if and only if \( tx \in \mathbb{C} \) for all \( t > 0 \). We say a function \( f : \mathbb{C} \to (0, \infty) \) is (multivariate) regularly varying with limit function \( \lambda \),
with $\lambda(x) > 0$ for $x \in \mathbb{C}$ and tail index $\beta$, if there exists a function $g : (0, \infty) \rightarrow (0, \infty)$ with $g \in \mathcal{RV}_\beta$ such that
\[ \lim_{t \to \infty} \frac{f(tx)}{g(t)} = \lambda(x), x \in \mathbb{C}; \] (4)
or equivalently, if there exists a function $b : (0, \infty) \rightarrow (0, \infty)$ with $b \in \mathcal{RV}_{-1/\beta}$
\[ \lim_{t \to \infty} tf(b(t)x) = \lambda(x), x \in \mathbb{C}. \] (5)

As a consequence, we have the homogeneity property that $\lambda(tx) = t^\beta \lambda(x), x \in \mathbb{C}, t > 0$. Moreover, we can obtain (5) from (4) by defining $b(t) = g^{-1}(1/t)$ where $f^{-1}$ denotes the generalized left inverse of a monotone function; since without loss of generality, $g$ can be assumed to be monotone; see Resnick (1987, 2007) for details. We write $f \in \mathcal{MRV}(\beta, g, \lambda, \mathbb{C})$ when following (4) or $f \in \mathcal{MRV}(\beta, b, \lambda, \mathbb{C})$ when following (5); one or more of the parameters are also often dropped for convenience. Both $b$ and $g$ are referred to as scaling functions.

2.3 Tail Behavior of the Univariate and Bivariate Graphex Marginals

The following result characterizes the asymptotic behaviour of $\mu_1$ and $\mu_2$ at infinity, given that $W$ is multivariate regularly varying. The proof uses ideas from [de Haan and Resnick (1987), Theorem 2.1]. A uniformity condition in addition to regular variation is imposed to guarantee regular variation of the integrals $\mu_1, \mu_2$, see [de Haan and Resnick (1987), Sect. 2] for further discussions. Let $\| \cdot \|$ denote a norm on $\mathbb{R}^2$.

**Theorem 1** Suppose the graphex function $W : \mathbb{R}^2_+ \rightarrow [0, 1]$ is regularly varying on $\mathbb{R}^2_+ \setminus \{0\}$ with limit function $\omega$ and scaling function $h \in \mathcal{RV}_\alpha$ where $\alpha > 1$. Moreover assume that $\omega$ is bounded on the unit sphere $\mathbb{S} := \{(x, y) \in \mathbb{R}^2_+ \setminus \{0\} : \|(x, y)\| = 1\}$ and satisfies the uniformity condition
\[ \lim_{t \to \infty} \sup_{(x, y) \in \mathbb{S}} \left| \frac{W(tx, ty)}{h(t)} - \omega(x, y) \right| = 0. \] (6)

Then the following hold:

(i) The univariate graphex marginal $\mu_1$ is regularly varying with tail index $-(\alpha - 1)$ and
\[ \lim_{t \to \infty} \frac{\mu_1(tx)}{t h(t)} = \int_0^\infty \omega(x, y) \, dy \quad x > 0. \] (7)

(ii) The bivariate graphex marginal $\mu_2$ is regularly varying on $(0, \infty)^2$ with tail index $-2(\alpha - 1)$ and
\[ \lim_{t \to \infty} \frac{\mu_2(tx, ty)}{t^2 h(t)} = \int_0^\infty \omega(x, z) \omega(y, z) \, dz, \] (8)
for $(x, y) \in (0, \infty)^2$.

(iii) The convergence in (8) is uniform on sets bounded away from the axes, i.e., for any fixed $\delta > 0$: 
\[
\lim_{t \to \infty} \sup_{x,y \geq 0} \left| \frac{\mu_2(tx, ty)}{t h^2(t)} - \int_0^\infty \omega(x, z) \omega(y, z) \, dz \right| = 0.
\]  

(9)

**Proof** First, using a change of variable \( v = tz \) we have

\[
\frac{\mu_1(tx)}{th(t)} = \int_0^\infty \frac{W(tx, z)}{th(t)} \, dz = \int_0^\infty \frac{W(tx, tv)}{h(t)} \, dv, \quad x > 0,
\]

(10)

\[
\frac{\mu_2(tx, ty)}{th^2(t)} = \int_0^\infty \frac{W(tx, z)W(ty, z)}{th^2(t)} \, dz = \int_0^\infty \frac{W(tx, tv)W(ty, tv)}{h(t)} \, dv,
\]

(11)

for \((x, y) \in (0, \infty)^2\). By our assumptions, the integrands in (10) and (11) are non-negative and converge to \(\omega(x, v)\) and \(\omega(x, v)\omega(y, v)\) respectively as \(t \to \infty\). Hence if we show that the integrands are bounded by an integrable function and the limits are integrable, then (7) and (8) hold using dominated convergence. Since all norms in \(\mathbb{R}^2\) are equivalent, without loss of generality let us fix a particular norm in \(\mathbb{R}^2\) given by \(\| (x, y) \| = |x| \vee |y|\).

(i) For fixed \(x > 0\), and any \(v \geq 0\), with \(t > 0\),

\[
\frac{W(tx, tv)}{h(t)} = \frac{W(t\| (x, v) \|, \frac{xv}{\| (x, v) \|})}{h(t\| (x, v) \|)} \frac{h(t\| (x, v) \|)}{h(t)},
\]

Since \(\| (x, v) \| \geq x > 0\), and \((x, v)/\| (x, v) \| \in \mathfrak{X}\), given a fixed \(\eta > 0\), we obtain from (6) that for \(t > t_0 \equiv t_0(x, \eta)\),

\[
\frac{W(t\| (x, v) \|, \frac{xv}{\| (x, v) \|})}{h(t\| (x, v) \|)} \leq \sup_{\alpha \in \mathfrak{X}} \omega(\alpha) + \eta < \infty,
\]

(12)

since \(\omega\) is bounded on \(\mathfrak{X}\). Using Potter’s bounds (Resnick 2007), for a large enough \(t\), \(h(t\| (x, v) \|)/h(t)\) is bounded by \(c\| (x, v) \|^{-\alpha + \rho}\) for a constant \(c > 0\) and \(\rho\) such that \(0 < \rho < \alpha - 1\). Hence,

\[
\frac{W(tx, tv)}{h(t)} \leq C\| (x, v) \|^{-\alpha + \rho},
\]

(13)

for some constant \(C > 0\), and

\[
\int_0^\infty C\| (x, v) \|^{-\alpha + \rho} \, dv = Cx^{-\alpha + \rho} \int_0^x dv + C \int_x^\infty u^{-\alpha + \rho} \, dv < \infty,
\]

(14)

which shows that the integrand in (10) is bounded by an integrable function. By homogeneity property we also have \(\omega(t(x, y)) = t^{-\alpha} \omega(x, y)\) for \((x, y) \in \mathbb{R}_+^2 \setminus \{0\}\) and hence

\[
\int_0^\infty \omega(x, y) \, dy = \int_0^\infty \| (x, y) \|^{-\alpha} \omega\left(\frac{(x, y)}{\| (x, y) \|}\right) \, dy.
\]

(15)
\[ \leq \sup_{\alpha \in \mathbb{R}} \omega(\alpha) \int_{0}^{\infty} (x \vee y)^{-\alpha} \, dy < \infty, \]  

(16)

since \( \omega \) is bounded in \( \mathbb{R} \) and \( \alpha > 1 \). Hence (7) is a consequence of \( W \in \mathcal{MRV}(-\alpha, h, \omega) \), (13), (14), (15) and the dominated convergence theorem. Therefore \( \mu_1 \in \mathcal{RV}_{-\alpha} \).

(ii) Using the bounds in (13), (14) (and finding similar bounds for \( W(t y, t v) / h(t) \)), we can conclude that for any \((x, y) \in (0, \infty) \times (0, \infty) \), the integrand in (11) is bounded by an integrable function. To see that the limit (as \( t \to \infty \)) of the integrand in (11) is integrable, observe that

\[ \int_{0}^{\infty} \omega(x, z) \omega(y, z) \, dz = \int_{0}^{\infty} \|(x, z)\|^{-\alpha} \omega\left(\frac{(x, z)}{\|(x, z)\|}\right) \|(y, z)\|^{-\alpha} \omega\left(\frac{(y, z)}{\|(y, z)\|}\right) \, dz \]

\[ \leq \left( \sup_{\alpha \in \mathbb{R}} \omega(\alpha) \right)^2 \int_{0}^{\infty} (x \vee z)^{-\alpha} (y \vee z)^{-\alpha} < \infty, \]

since \( \omega \) is bounded in \( \mathbb{R} \) and \( \alpha > 1 \). Then (11) follows from the dominated convergence theorem, and we have

\[ \mu_2 \in \mathcal{MRV}\left(-2 - 1, th^2(t), \int_{0}^{\infty} \omega(x, z) \omega(y, z) \, dz, (0, \infty)^2\right). \]

(iii) From [de Haan and Resnick (1987), Theorem 2.1] we know that the uniformity condition (6) also implies that for any \( \delta > 0 \),

\[ \lim_{t \to \infty} \sup_{\|(x, y)\| \geq \delta} \left| \frac{W(tx, ty)}{h(t)} - \omega(x, y) \right| = 0. \]  

(17)

Now for fixed \( \delta > 0 \) and \( x \land y > \delta \),

\[
\left| \frac{\mu_2(tx, ty)}{th^2(t)} - \int_{0}^{\infty} \omega(x, z) \omega(y, z) \, dz \right| \leq \int_{0}^{\infty} \left| \frac{W(tx, tz)}{h(t)} - \omega(x, z) \omega(y, z) \right| \, dz
\]

\[ \leq \int_{0}^{\infty} \left| \frac{W(tx, tz)}{h(t)} - \omega(y, z) \right| \, dz
\]

\[ + \int_{0}^{\infty} \omega(y, z) \left| \frac{W(tx, tz)}{h(t)} - \omega(x, z) \right| \, dz. \]

Since \( \{(x, y) \in \mathbb{R}_+^2 : x \land y > \delta \} \subset \{(x, y) \in \mathbb{R}_+^2 : x \lor y = \|(x, y)\| > \delta \} \), then by (17), given \( \epsilon > 0 \), there exists \( t_0 \) such that for \( t > t_0 \), both \( \left| \frac{W(tx, tz)}{h(t)} - \omega(y, z) \right| < \epsilon \) and \( \left| \frac{W(tx, tz)}{h(t)} - \omega(x, z) \right| < \epsilon \) for any \((x, y)\) with \( x \land y > \delta \). Therefore for \( t > t_0 \),

\[
\sup_{x/y > \delta} \left| \frac{\mu_2(tx, ty)}{th^2(t)} - \int_{0}^{\infty} \omega(x, z) \omega(y, z) \, dz \right| \leq \epsilon \sup_{x > \delta} \int_{0}^{\infty} \frac{W(tx, tz)}{h(t)} \, dz
\]

\[ + \epsilon \sup_{y > \delta} \int_{0}^{\infty} \omega(y, z) \, dz = \epsilon \sup_{x > \delta} A_{x,t} + \epsilon \sup_{y > \delta} B_{y} \quad (say). \]
From (13) and (14), since \( x > \delta > 0 \), for large enough \( t \), \( \sup_{x,t} A_{x,t} \leq C_1 \) for some constant \( C_1 > 0 \). Also using (15), we can check that \( \sup_{y > \delta} B_y \leq C_2 \) for some constant \( C_2 > 0 \). Therefore

\[
\lim_{t \to \infty} \sup_{x,y \geq \delta} \left| \frac{\mu_2(tx, ty)}{th^2(t)} - \int_0^\infty \omega(x, z) \omega(y, z) \, dz \right| \leq \epsilon (C_1 + C_2)
\]

and since \( \epsilon > 0 \) is arbitrary, we have (9).

\[ \square \]

**Example 1** Consider the graphex function \( W : \mathbb{R}_+^2 \to (0, \infty) \) with \( \alpha > 1 \):

\[
W(x, y) = \frac{1}{1 + x^\alpha + y^\alpha}.
\]  

(18)

Here

\[
W \in \mathcal{MRV}(-\alpha, h(t) = t^{-\alpha}, \omega(x, y) = \frac{1}{x^\alpha + y^\alpha}, (0, \infty)^2).
\]

and we can check that the uniformity condition (6) holds. Therefore the univariate marginal \( \mu_1 \in \mathcal{RV}_{(-\alpha, -1]} \) and for \( x > 0 \),

\[
\lim_{t \to \infty} \frac{\mu_1(tx)}{th(t)} = \int_0^\infty \omega(x, y) \, dy = x^{\alpha+1} \int_0^\infty \frac{1}{1 + z^\alpha} \, dz = \frac{\pi \cosec(\pi \alpha)}{x^{\alpha+1}}.
\]

See [Gradshteyn and Ryzhik (2007), S3.241.2]. Moreover, the bivariate marginal satisfies

\[
\mu_2 \in \mathcal{MRV}(-2\alpha - 1, th^2(t), \nu(\cdot), (0, \infty)^2),
\]

where \( \nu \) is hard to compute in closed form in general. For \( \alpha = 2 \), we can compute \( \nu \) in closed form, which is given by

\[
\nu(x, y) = \lim_{t \to \infty} \frac{\mu_2(tx, ty)}{th^2(t)} = \int_0^\infty \frac{1}{x^2 + z^2} \frac{1}{y^2 + z^2} \, dz = \frac{\pi}{2x^2y + 2xy^2},
\]

for \( (x, y) \in (0, \infty)^2 \).

**Remark 1** A classical example of a multivariate regularly varying \( W \) is when it is *separable*, i.e., \( W(x, y) = U(x)U(y) \) with \( U \in \mathcal{RV}_{-\alpha} \) for some \( \alpha > 1 \). Unfortunately, the uniformity condition (6) fails to hold in this case and we cannot apply Theorem 1. Nevertheless, we can still ascertain the tail behavior in this case easily because of the separable structure.

**Lemma 1** Let \( W : \mathbb{R}_+^2 \to [0, 1] \) be a graphex function such that \( W(x, y) = U(x)U(y) \) where \( U : \mathbb{R}_+ \to (0, \infty) \) satisfies \( U \in \mathcal{RV}_{-\alpha} \) for some \( \alpha > 1 \). Then the following holds:

1. \( W \in \mathcal{MRV}(-2\alpha, U^2(t), \omega(x, y), (0, \infty)^2) \) where

\[ \square \]
\[
\lim_{t \to \infty} \frac{W(tx, ty)}{U^2(t)} = x^{-\alpha}y^{-\alpha} = : \omega(x, y), \quad (x, y) \in (0, \infty)^2,
\]

(ii) \( \mu_1 \in \mathcal{RV}_{-\alpha} \) where
\[
\lim_{t \to \infty} \frac{\mu_1(tx)}{U(t)} = x^{-\alpha} \int_0^\infty U(z) \, dz, \quad x > 0,
\]

(iii) \( \mu_2 \in \mathcal{MRV}(-2\alpha, U^2(t), C\omega(x, y), (0, \infty)^2) \) where \( C = \int_0^\infty U^2(z) \, dz \).

**Proof** The proof is an easy consequence of \( U \in \mathcal{RV}_{-\alpha} \) and is omitted here.

\[ \square \]

**Example 2** Consider the graphex function \( W : \mathbb{R}^2_+ \to (0, \infty) \) with \( \alpha > 1 \):
\[
W(x, y) = U(x)U(y) := \frac{1}{1 + x^\alpha} \frac{1}{1 + y^\alpha}.
\]
Clearly,
\[
W \in \mathcal{MRV}(-2\alpha, U^2(t), \omega(x, y) = \frac{1}{x^\alpha y^\alpha}, (0, \infty)^2).
\]
Using Example 1, the univariate marginal \( \mu_1 \in \mathcal{RV}_{-\alpha} \) and for \( x > 0 \),
\[
\lim_{t \to \infty} \frac{\mu_1(tx)}{U(t)} = x^{-\alpha} \int_0^\infty U(z) \, dz = x^{-\alpha} \int_0^\infty \frac{1}{1 + z^\alpha} \, dz = \frac{\pi}{\alpha} \text{cosec}(\pi/\alpha) x^{-\alpha}.
\]
Moreover,
\[
\mu_2 \in \mathcal{MRV}(-2\alpha, U^2(t), C\omega(x, y), (0, \infty)^2),
\]
where
\[
C = ((\pi/\alpha)\text{cosec}(\pi/\alpha))^2.
\]

3 Common Connections

In this section, our goal is to understand the behavior of the number of common connections between two vertices, first for a fixed pair of vertices and eventually for a randomly chosen pair in a graph driven by the (Kallenberg exchangeable) graphex process as described in Sect 2.

3.1 Distribution of Common Connections in \( \mathcal{G}_t \)

Recall that from the unit-rate Poisson process \( \Pi = (\theta_i, \eta_i)_{i=1,2,\ldots} \), the finite-size graphex \( \mathcal{G}_t \) is created by restricting \( \sum_{i,j} Z_{ij} \delta_{(\theta_i, \eta_j)} \) to \([0, t]^2\), where \( Z_{ij} \sim \text{Bernoulli}(W(\eta_i, \eta_j)) \), and \( Z_{ij} = Z_{ji} \). The first co-ordinate in \( \Pi \) does not provide any relevant information other than
picking the plausible vertices, so for ease of computation, we project \( I \) to its second coordinate as

\[
\Pi^\eta = \{ \eta_i | (\theta_i, \eta_i) \in \Pi_i \},
\]

so that \( \Pi^\eta \) is a one-dimensional Poisson process with rate \( t \).

Let \( I \) be an index set. Then following [Veitch and Roy (2015), Sect. 5], for a locally finite, simple sequence \( \phi = (x_i)_{i \in I} \), and a sequence of values \( \{u_{ij}\} \in [0,1] \) with \( u_{ij} = u_{ji}, i,j \in I \), we define the 2nd order common connections function or 2-c-degree function at (vertices) \( x_i, x_j \in \phi \) to be

\[
C(x_i, x_j, \phi, (u_{ij})) := \sum_{x_k \in \Pi^\eta_{i \cup j}} \mathbb{I}_{\{\{w(x_i, x_k) \geq u_{ik}\} \cap \{w(x_j, x_k) \geq u_{jk}\}\}}.
\]

(19)

With this definition, for the graphex \( G_t \), realized at \( Z_{ij} = z_{ij} \) (which are realizations of Uniform(0, 1) random variables), the 2-c-degree function at (\( \theta_1, \eta_1 \), \( \theta_2, \eta_2 \) \( ) \in \Pi_i \) is

\[
C(\eta_1, \eta_2, \Pi^\eta_i, (z_{ij})).
\]

(20)

Taking \( \eta_1, \eta_2 \) to be the proxy for the points \( (\theta_1, \eta_1), (\theta_2, \eta_2) \in \Pi_i \), we refer to \( C(\eta_1, \eta_2, \Pi^\eta_i, (z_{ij})) \) as the number of common friends/connections of \( \eta_1, \eta_2 \) in \( G_t \). We can analogously define a \( k \)-c-degree function, for the number of common connections between \( k \geq 2 \) vertices as well.

We now explore how \( C(\eta_1, \eta_2, \Pi^\eta_i, (Z_{ij})) \) behaves for fixed \( t \), and later in Sect. 3.2, we study the asymptotic behavior of common friends for a pair of randomly chosen vertices in \( G_t \), as \( t \to \infty \). Since for a finite set of points \( x_1, \ldots, x_k \in \mathbb{R}_+ \), the probability that \( x_1, \ldots, x_k \in \Pi^\eta_i \) is 0, it is traditional to use some tools from Palm theory. First note that the conditional distribution of \( \Pi^\eta_i \) given \( x_1, \ldots, x_k \) is equivalent to the distribution of \( \Pi^\eta_i \cup \{x_1, \ldots, x_k\} \) which is the (extended) Slivnyak-Mecke theorem if \( \Pi^\eta_i \) is a Poisson process; see [Moller and Waagepetersen (2003), Theorem 3.3]. To make the paper self-contained, we state the extended Slivnyak-Mecke theorem below.

**Theorem 2** Let \( S \subset \mathbb{R}^d \) and define \( \mathcal{N} \) as the set of locally finite subsets of \( S \). Let \( X = (x_i)_{i \geq 1} \in \mathcal{N} \) be a Poisson point process on \( S \) with a diffuse and locally finite mean measure \( \mu \). Then, for any \( n \geq 1 \) and any measurable function \( h : \mathcal{N} \times S^n \mapsto [0, \infty) \),

\[
\mathbb{E} \left[ \sum_{\xi_1, \ldots, \xi_n \in X} h(X \setminus \{\xi_1, \ldots, \xi_n\}) \right] = \int_{S^n} \mathbb{E}[h(X \setminus \{\xi_1, \ldots, \xi_n\})] \mu(d\xi_1) \cdots \mu(d\xi_n).
\]

With Theorem 2 available, we present in the following proposition that the 2-c-degree of two fixed vertices in a graphex \( G_t \) is Poisson distributed.

**Proposition 1** Let \( x, y \in \mathbb{R}_+ \), and \( \zeta_{(i,j)} \) be a symmetric array of iid Uniform [0, 1] random variables. Then \( C(x,y, \Pi^\eta_i \cup \{x,y\}, (\zeta_{(i,j)})) \) follows a Poisson distribution with rate \( t\mu_2(x,y) \).

**Proof** In the sequel, we suppress the notation as

\[
C(x,y) \equiv C(x,y, \Pi^\eta_i \cup \{x,y\}, (\zeta_{(i,j)})),
\]

and use \( i(s) \equiv i(s, \Pi^\eta_i) \) to denote the index of the point \( s \in \Pi^\eta_i \) with respect to the natural ordering on \( \mathbb{R}_+ \). Then we have
Note also that
\[ C(x, y) = \sum_{z \in \Pi_t} \mathbb{1}_{\{W(x, z) \geq \zeta_t(1/2)\}} \cdot \mathbb{1}_{\{W(y, z) \geq \zeta_t(1/2)\}}. \]

Then by Campbell’s theorem (cf. [Kingman (1993), Sect. 5.3]), the characteristic function of \( C(x, y) \) is

\[ \mathbb{E}[\exp(itC(x, y))] = \mathbb{E}[\exp(it \sum_{z \in \Pi_t} \mathbb{1}_{\{W(x, z) \geq \zeta_t(1/2)\}} \cdot \mathbb{1}_{\{W(y, z) \geq \zeta_t(1/2)\})}] \]

\[ = \exp \left\{ \int_{\mathbb{R}_+} \int_{[0,1]} \int_{[0,1]} (1 - e^{it \mathbb{1}_{\{u \leq W(x, z)\}} \mathbb{1}_{\{v \leq W(y, z)\}}}) t du dv dz \right\} \]

\[ = \exp \left\{ \sum_{n=1}^{\infty} \frac{(it)^n}{n!} \int_{\mathbb{R}_+} \int_{[0,1]} \int_{[0,1]} \mathbb{1}_{\{u \leq W(x, z)\}} \mathbb{1}_{\{v \leq W(y, z)\}} du dv dz \right\} \]

\[ = \exp \left\{ t \mu_2(x, y)(e^{it} - 1) \right\}. \]

Hence, \( C(x, y) \) is a Poisson random variable with rate \( t \mu_2(x, y) \).

Proposition 1 characterizes the distribution for a fixed pair of \((x, y)\) in \(G_t\) with \(t\) also fixed. In the next section, we extend our analysis to the asymptotic behavior of the 2-c-degree distribution when \(t\) goes to infinity.

### 3.2 Asymptotic Distribution of Common Connections

To analyze the asymptotic distribution, we start with a simulated example where \( W(x, y) = (1 + x + y)^{-3}, \ x, y \geq 0 \). We compute the empirical distribution of the number of common friends, i.e.,

\[ \sum_{x, y \in \Pi_t} \mathbb{1}_{\{C(x, y) = k\}} \sum_{x, y \in \Pi_t} \mathbb{1}_{\{C(x, y) > 0\}}, \quad k \geq 1, \quad \text{(21)} \]

and numerical results are plotted in Fig. 1. We observe that the plot looks linear until a certain point and then it quickly tapers down, indicating a certain cut-off prior to which a power-law behavior is quite evident. Excluding data beyond the cut-off, we estimate the slope of the log-log plot and obtain an estimated slope between 1 and 2. From this we hypothesize that the 2-c-degree of a randomly chosen pair of vertices, \( D_c \), is roughly following

\[ \mathbb{P}(D_c = k) \approx k^{-1-a}, \quad a \in (0, 1), \]
for large $k$, which implies, for $t$ large,

$$t\mathbb{P}\left(\frac{D}{t^{1/a}} > x\right) \approx x^{-a}. \quad (22)$$

Since $1/a > 1$, then for a randomly chosen pair of vertices (as long as we move away from the cut-off), the growth rate of the common friends is faster than $O(t)$. Note that by the construction of the model, for $x, y$ small, the expected number of common friends between vertices with labels $x, y$ is $t\mu_2(x, y)$, which is growing no faster than $O(t)$. This leads to the cut-off as observed in the figure. Later in Theorem 3, we stay away from the cut-off, and give the theoretical analysis on the asymptotic features for $x, y \geq b(t)e$, with $b(t) \to \infty$.

**Definition 3** Define $N'_c(k)$ as the number of vertex pairs with second co-ordinate value $\eta > b(t)e$ and who have $k \geq 1$ common friends in the graph $G_t$, i.e.

$$N'_c(k) := \sum_{x,y \in \Pi^e, x,y > b(t)e} \mathbb{1}_{\{C(x,y) = k\}}, \quad e > 0.$$

We study the behavior of $N'_c(k)$ when $W$ is multivariate regularly varying. When $W \in \mathcal{MRV}(-\alpha, h(t), \omega(x,y), \mathbb{R}^2_+ \setminus \{0\})$ is non-separable, we choose the scaling function to be $b \in \mathcal{RV}_{1/(2\alpha-1)}$, $\alpha > 2$, such that

$$\lim_{t \to \infty} t\mu_2(b(t)x, b(t)y) = \int_0^\infty \omega(x,z)\omega(y,z)dz. \quad (23)$$

When $W$ is separable, i.e., $W(x,y) = U(x)U(y)$, and $W \in \mathcal{MRV}(-2\alpha, U^2(t), \omega(x,y), (0, \infty)^2)$, we set $b(t) = (1/U)^{-\sqrt{t}}$, so that $b \in \mathcal{RV}_{1/(2\alpha)}$, $\alpha > 1$ and

$$\mu_2 \in \mathcal{MRV}\left(-2\alpha, b(t), \int_0^\infty \omega(x,z)\omega(y,z)dz, (0, \infty)^2\right).$$

Assume that for $q > 1/2$ and some constant $C_4 > 0$, we have
\[ \mu_4(x_1, x_2, x_3, x_4) \leq C_4 \left( \mu_2(x_1, x_2) \mu_2(x_3, x_4) \right)^q. \]  

(24)

We now give the asymptotic behavior of \( \mathbb{E}[N^c_t(k)] \) for both separable and non-separable cases.

**Theorem 3** Suppose (i) \( W \in \mathcal{MRV}(-\alpha, \mathbb{R}^2 \setminus \{0\}) \) and condition (6) is satisfied, or, (ii) \( W(x, y) = U(x)U(y) \) with \( U \in \mathcal{RV}_{-\alpha} \) for \( \alpha > 0 \). Also assume that (24) is satisfied, and that \( N^c_t(k) \) is as defined in Definition 3. Let \( b(t) \in \mathcal{RV}_{1/\gamma} \) for \( \gamma > 1 \) and \( \lambda(x, y) \) be some functions satisfying

\[
\lim_{t \to \infty} \mu_2(b(t)x, b(t)y) = \lambda(x, y), \quad x, y > 0.
\]

(25)

Then for \( k \geq 1 \), we have

\[
\frac{N^c_t(k)}{(tb(t))^2} \to \frac{1}{k!} \int_\mathbb{R} \int_\mathbb{R} (\lambda(u, v))^k e^{-\lambda(u,v)} \, du \, dv.
\]

(26)

**Proof** First note that (25) is a consequence of either Theorem 1 where \( \gamma = 2\alpha - 1 \), or, Lemma 1 where \( \gamma = 2\alpha \). From Proposition 1, we have for \( x, y > 0 \), \( C(x, y) \) follows a Poisson distribution with rate \( \mu_2(x, y) \). Now by the extended Slivnyak-Mecke theorem given in Theorem 2, we see that

\[
\mathbb{E}[N^c_t(k)] = t^2 \int_{b(t)e}^{\infty} \int_{b(t)e}^{\infty} \mathbb{P}[C(x, y) = k] \, dx \, dy
\]

\[
= t^2 \int_{b(t)e}^{\infty} \int_{b(t)e}^{\infty} \frac{(\mu_2(x, y))^k}{k!} \, e^{-\mu_2(x,y)} \, dx \, dy
\]

\[
= t^2 b(t)^2 \int_\mathbb{R} \int_\mathbb{R} \frac{(\mu_2(b(t)u, b(t)v))^k}{k!} \, e^{-\mu_2(b(t)u,b(t)v)} \, du \, dv,
\]

(27)

which, either by using Theorem 1, or, Lemma 1, converges to the limit on the right hand side of (26) when divided by \( t^2 b(t)^2 \) (as \( t \to \infty \)). With (27) available, we then show the convergence in (26) by proving

\[
\text{Var}[N^c_t(k)] = O \left( t^{-\kappa} \mathcal{C}(t) \mathbb{E}[N^c_t(k)]^2 \right),
\]

(28)

for some \( \kappa > 0 \) and a slowly varying function \( \mathcal{C}(t) \). Note that

\[
\left( N^c_t(k) \right)^2 = N^c_t(k)
\]

\[+ \sum_{x_i \in \Pi_1, i=1,2,3} \mathbb{I}[C(x_1, x_2) = k, C(x_1, x_3) = k] \mathbb{I}[x_1 \neq x_2 \neq x_3] \mathbb{I}[x_i > b(t)c, i=1,2,3] \]

\[+ \sum_{x_i \in \Pi_1, i=1,\ldots,4} \mathbb{I}[C(x_1, x_2) = k, C(x_1, x_3) = k] \mathbb{I}[x_1 \neq x_2 \neq x_3 \neq x_4] \mathbb{I}[x_i > b(t)c, i=1,\ldots,4] \]

\[=: N^c_t(k) + A^c_t(k) + B^c_t(k). \]

Since

\[
\mu_4(x_1, x_2, x_3, x_4) \leq \frac{1}{2} \left( \mu_2(x_1, x_2) + \mu_2(x_3, x_4) \right),
\]
then applying Theorem 2 and the assumption in (24) gives that

\[
\mathbb{E}[B_t'(k)] - (\mathcal{E}[N_t'(k)])^2 \leq t^1 \sum_{l=0}^{k-1} \int_{(b(t)\epsilon, \infty)^3} \frac{(t\mu_2(x_1, x_2))^l}{l!} \frac{(t\mu_2(x_3, x_4))^l}{l!} \frac{(t\mu_4(x_1, x_2, x_3, x_4))^{k-l}}{(k-l)!} dx_1 dx_2 dx_3 dx_4 \\
\times e^{-t\mu_2(x_1, x_2) + \mu_2(x_3, x_4) - \mu_4(x_1, x_2, x_3, x_4)} dx_1 dx_2 dx_3 dx_4 \\
\leq t^1 \sum_{l=0}^{k-1} \int_{(b(t)\epsilon, \infty)^3} \frac{(t\mu_2(x_1, x_2))^l}{l!} \frac{(t\mu_2(x_3, x_4))^l}{l!} \frac{(t\mu_4(x_1, x_2, x_3, x_4))^{k-l}}{(k-l)!} dx_1 dx_2 dx_3 dx_4 \\
\times e^{-t/2(\mu_2(x_1, x_2)+\mu_2(x_3, x_4))} dx_1 dx_2 dx_3 dx_4 \\
\leq \sum_{l=0}^{k-1} C^2_4 t^{4+k+l} (l!)^2 (k-l)! \left( \int_{(b(t)\epsilon, \infty)^3} (\mu_2(x_1, x_2))^{l+q(k-l)} e^{-t/2(\mu_2(x_1, x_2))} dx_1 dx_2 \right)^2,
\]

and using the results in (27), the upper bound above is of order

\[
O(t^{4-(2q-1)(b(t))^4}) = O(t^{4-(1/c)}(b(t))^4).
\]

For \(\mu_3\), we have that

\[
\mu_3(x_1, x_2, x_3) = \int_0^{\infty} \prod_{i=1}^3 W(x_i, z) dz \\
= \int_0^{\infty} \left( \sqrt{W(x_1, z)W(x_2, z)} \right) \left( \sqrt{W(x_1, z)W(x_3, z)} \right) dz \\
\leq \left( \int_0^{\infty} W(x_1, z)W(x_2, z) dz \right)^{1/2} \left( \int_0^{\infty} W(x_1, z)W(x_3, z) dz \right)^{1/2} \\
\leq \sqrt{\mu_2(x_1, x_2)\mu_2(x_1, x_3)}.
\]

Then, similarly, for \(\mathbb{E}[A_t^*(k)]\), the extended Slivnyak-Mecke theorem gives:

\[
\mathbb{E}[A_t^*(k)] \leq t^3 \sum_{l=0}^{k} \int_{(b(t)\epsilon, \infty)^3} \frac{(t\mu_2(x_1, x_2))^l}{l!} \frac{(t\mu_2(x_1, x_3))^l}{l!} \frac{(t\mu_3(x_1, x_2, x_3))^{k-l}}{(k-l)!} dx_1 dx_2 dx_3 dx_4 dx_5 dx_6 \\
\times e^{-t(\mu_2(x_1, x_2)+\mu_2(x_1, x_3)+\mu_3(x_1, x_2, x_3))} dx_1 dx_2 dx_3 dx_4 dx_5 dx_6 \\
\leq \sum_{l=0}^{k} \frac{t^{3+k+l}}{(l!)^2 (k-l)!} \int_{(b(t)\epsilon, \infty)^3} (\mu_2(x_1, x_2))^{l+(k-l)/2} (\mu_2(x_1, x_3))^{l+(k-l)/2} \times \times e^{-t/2(\mu_2(x_1, x_2)+\mu_2(x_1, x_3))} dx_1 dx_2 dx_3,
\]

and applying the same argument as in (27) the bound is of order

\[
O(t^3(b(t))^3) = O(t^{(1+1/\gamma)}\mathcal{E}(t)(b(t))^4),
\]

where \(b(t) \in K\mathcal{V}_{1/\gamma}\) and we write \(1/b(t) = t^{-(1+1/\gamma)}\mathcal{E}(t)\) for some slowly varying function \(\mathcal{E}(t)\). Combining (31) with (30), (28) holds with \(\kappa \equiv \min\left(2q-1, 1 + \frac{1}{\gamma}\right) > 0\).
Remark 2 When $W(\cdot, \cdot)$ is separable, the condition in (24) holds with $q = 1$. Then the upper bounds in (29) and (31) are of orders
\[ O\left(t^{2/a+3}\left(\ell'_1(t)\right)^4\right), \quad O\left(t^{3/(2a)+3}\left(\ell'_1(t)\right)^3\right), \]
respectively, where $\ell'_1(t)$ is a slowly varying function such that $b(t) = t^{1/(2a)}\ell'_1(t)$. Therefore, the convergence results in Theorem 3 hold for $W(\cdot, \cdot)$ separable.

3.2.1 Tail Asymptotics for Common Connections

Note that the convergence results in Theorem 3 do not provide an explicit explanation for the power-law behavior as observed in Fig. 1. Hence, we proceed with a detailed discussion on the tail distribution of 2-c-degrees. Consider the function
\[ f(x) = x^ke^{-x}, \quad x \in A, \]
and we see that $f$ is increasing in $x$ if $k \geq \sup_{x \in A} x$.

Separable case In the separable case, assume in addition that $U(x)U(y) \leq U(x+y)$, $x, y \geq 0$, then the integrand in (27) satisfies
\[ \left(t\mu_2(b(t)x, b(t)y)\right)^ke^{-t\mu_2(b(t)x, b(t)y)} \leq (C \cdot tU(b(t)(x+y)))^ke^{CtU(b(t)(x+y))}, \]
with $C = \int_0^\infty U^2(z)dz$, as long as $k \geq \sup_{x,y \geq b(t)\epsilon} t\mu_2(b(t)x, b(t)y) = \lambda(\epsilon, \epsilon)$. Therefore, for $k \geq \lambda(\epsilon, \epsilon)$, we have
\[ \frac{1}{k!} \int_\epsilon^\infty \int_\epsilon^\infty \left(t\mu_2(b(t)x, b(t)y)\right)^ke^{-t\mu_2(b(t)x, b(t)y)}dxdy \leq \frac{1}{k!} \int_\epsilon^\infty \int_\epsilon^\infty \left(C \cdot tU(b(t)(x+y))\right)^ke^{-CtU(b(t)(x+y))}dxdy, \]
and since $b(t) = (1/U)^{-\sqrt{t}}$ in the separable case, then for some slowly varying function $\ell'$, we have
\[ tU(b(t)(x+y)) = t^{1/2}(x+y)^{-a} \frac{\ell'(b(t)(x+y))}{\ell'(b(t))}. \]
Assume further that
\[ \lim_{t \to \infty} \ell'(t) = C_0 > 0, \]
then the upper bound in (32) is of order $O(t^{1/a}k^{-1-2/a})$. By (22) we may write $k = O(t^{1/a})$, $a \in (0, 1)$, then the upper bound is of order
\[ k^{-\frac{1}{a}(2-a)-1}, \]
and $\frac{1}{a}(2-a) + 1 \in (1 + 1/a, 1 + 2/a)$.

Next, we give a lower bound for the asymptotic power-law behavior of the c-degrees in the separable case. Since
\[
\mu_2(x, y) = CU(x)U(y) \leq \frac{C}{2} (U^2(x) + U^2(y)),
\]
then we have for \( b \in \mathcal{R}\mathcal{V}_{1/(2a)}^+ \)
\[
\frac{b^2(t)}{k!} \int_{e}^{\infty} \int_{e}^{\infty} \left( t\mu_2(x, y) \right)^k e^{-t\mu_2(x, y)} \, dx \, dy
\]
\[
\geq \frac{1}{k!} \int_{e}^{\infty} \int_{e}^{\infty} \left[ tCU(b(t)x)U(b(t)y) \right]^k e^{-t\frac{c}{2}U^2(b(t)x) + U^2(b(t)y)} \, dx \, dy
\]
\[
= C^{\frac{k}{k!}} \left[ \int_{0}^{\infty} U^k(b(t)x) e^{-\frac{c}{2}U^2(b(t)x)} \, dx - \int_{0}^{e} U^k(b(t)x) e^{-\frac{c}{2}U^2(b(t)x)} \, dx \right]^2.
\]
Using [Caron et al. (2020), Lemma S3.5], since \( U^2(x) \in \mathcal{R}\mathcal{V}_{-2a}^+ \)
\[
\int_{0}^{\infty} U^k(x) e^{-\frac{c}{2}U^2(x)} \, dx \sim \frac{1}{2\alpha} \left( \frac{tC}{2} \right)^{-\frac{1}{2} - \frac{1}{2\alpha}} \Gamma \left( \frac{k}{2} - \frac{1}{2\alpha} \right) \ell'_1(t),
\]
for some slowly varying function \( \ell'_1(t) \). Note that when \( \epsilon \) is small,
\[
\int_{0}^{e} \left( t^{1/2} U(b(t)x) \right)^k e^{-\frac{c}{2}U^2(b(t)x)} \, dx \approx \int_{0}^{e} x^{-ak} e^{-\frac{c}{2}x^{2a}} \, dx,
\]
which is also small. Therefore, using Stirling’s approximation,
\[
\frac{1}{k!} \int_{b(t)c}^{\infty} \int_{b(t)c}^{\infty} \left( t\mu_2(x, y) \right)^k e^{-t\mu_2(x, y)} \, dx \, dy \geq C^{\frac{k}{k!}} \left( \frac{tC}{2} \right)^{-\frac{1}{2} - \frac{1}{2k}} \Gamma \left( \frac{k}{2} - \frac{1}{2k} \right) \ell'_1(t)^2
\]
\[
= \frac{C^{1/a}}{2^{2+1/a-1/2k}k^2} \cdot t^{\frac{1}{2} + \frac{1}{2\alpha}} \cdot \left( \frac{\Gamma \left( \frac{k}{2} - \frac{1}{2\alpha} \right)}{\Gamma \left( \frac{k}{2} + 1 \right)} \right)^2 \left( \frac{\Gamma \left( \frac{k}{2} + 1 \right)}{\Gamma(k+1)} \right)^2 \ell'_1(t)^2
\]
\[
\approx \frac{C^{1/a}}{2^{2+1/a-1/2k}k^2} \cdot t^{\frac{1}{2} + \frac{1}{2\alpha}} \cdot \left( \frac{k}{2} \right)^{-\frac{1}{2} - \frac{1}{2k}} \cdot \frac{\sqrt{2\pi k}}{2^{k+1}} \ell'_1(t)^2
\]
\[
= O \left( t^{\frac{1}{2} + \left( \frac{1}{2} + \frac{1}{2k} \right)} \right),
\]
if we assume (33). Hence for small \( \epsilon \) and \( k \) is large, (34) provides a lower bound for the asymptotic power-law behavior of the 2-c-degree distribution, i.e., for large \( t \),
\[
\frac{N_{i}^{c}(k)}{t^2b(t)^2} \approx O \left( k^{-\left( \frac{1}{2} + \frac{1}{2k} \right)} \right) \quad \text{(in probability)}.
\]

**Non-separable case** In the non-separable case, we need the following technical assumption in order to further study the limit behavior of \( \mathbb{E}[N_{i}^{c}(k)]/(tb(t))^2 \). For \((x, y) \in (0, \infty)^2\), assume there exist constants \( C_0 > 0 \) and \( x_0 \geq 0 \) such that for all \( x, y \geq x_0 \),
\[
W(x, y) \leq C_0 (\mu_1(x))^\theta (\mu_1(y))^\theta, \quad \mu_1(x_0) > 0,
\]
where \( \theta > 1/2 \) is a positive constant. As discussed in Caron et al. (2020), the assumption in (35) is satisfied with \( \theta = 0 \) when the function \( W \) is separable. Another example is
\[ W(x, y) = (1 + x + y)^{-\alpha}, \quad \alpha > 2, \] (36)

then we have
\[ \mu_1(x) = \frac{1}{\alpha - 1} (1 + x)^{-(\alpha - 1)}, \quad W(x, y) \leq (\alpha - 1)^{\frac{\alpha}{\alpha - 1}} (\mu_1(x) \mu_1(y))^{\frac{\alpha}{\alpha - 1}}. \]

In fact, the assumption in (35) implies for \( x, y > 0, \)
\[ \mu_2(x, y) = \int_0^\infty W(x, z)W(y, z)dz \leq C_0^2 \int_0^\infty \mu_1(z)^{2\theta}dz \cdot (\mu_1(x)\mu_1(y))^\theta =: C'(\mu_1(x)\mu_1(y))^\theta. \]

Let \( b \in \mathcal{R}V_{1/(2\alpha - 1)} \) be some scaling function such that (23) holds. Assume further that \( \mu_1(x)\mu_1(y) \leq \mu_1(x + y) \) then we have for \( x, y \geq \epsilon > 0, \)
\[ t\mu_2(b(t)x, b(t)y) \leq t(\mu_1(b(t)(x + y)))^\theta \]
\[ = t^{1 - \frac{\alpha(a-1)}{2\alpha - 1} (x + y)^{-(a-1)}} \frac{\ell_0(b(t)(x + y))}{(\ell_b(t))^{a-1}}, \]

where \( \ell_0(\cdot), \ell_b(\cdot) \) are two slowly varying functions satisfying
\[ \mu_1(x) = \epsilon^{-(a-1)} \ell_0(x), \quad b(t) = t^{1/(2\alpha - 1)} \ell_b(t). \]

Suppose also both \( \ell_0 \) and \( \ell_b \) satisfy (33). Then repeating the reasoning in the separable case, we see that the upper bound for
\[ \frac{1}{\epsilon} \int_\epsilon^\infty \int_\epsilon^\infty \left(t\mu_2(b(t)x, b(t)y)\right)^k e^{-t\mu_2(b(t)x, b(t)y)} dx dy \]
is of order
\[ \frac{1 - \frac{2}{(\alpha-1)}}{t^{\frac{\alpha(a-1)}{2\alpha - 1}}} k^{-1 - \frac{2}{\alpha(a-1)}} = t^{\frac{a}{2\alpha - 1}} k^{-1 - \frac{4}{a}}. \]

Recall (22), and we may write \( k = O(t^{1/a}) \), with \( a \in (0, 1) \), then the upper bound is of order
\[ k^{-1 - \frac{4}{a}(1-a)} - \frac{2a}{2a-1}, \]

and \( 1 + \frac{4}{a}(1-a) + \frac{2a}{2a-1} \in (1 + 2/(2\alpha - 1), 1 + 4/a) \). Note that for the example in (36), we have \( \ell_0, \ell_b \equiv 1 \) and \( \theta = \frac{a}{2(a-1)} \), which satisfy all those conditions listed above.

## 4 Simulation Studies

To see how tight the bounds derived in the previous section are, especially when the entire network is considered (e.g. the scenario in Fig. 1), we further examine the slope estimates through simulation studies. Since the major focus in this section is to check the tightness of bounds, we restrict our simulation experiments to two specific choices of \( W: \)
(1) Separable case: \( W(x, y) = (1 + x)^{-\alpha}(1 + y)^{-\alpha} \).
(2) Non-separable case: \( W(x, y) = (1 + x + y)^{-\alpha} \).

Recall also the Criterion (a) in Definition 1, and we note that we must set \( \alpha > 1 \) in the separable case and \( \alpha > 2 \) in the non-separable case to make Definition 1(a) satisfied.

We now choose two different sets of values for \( \alpha \):

(1) Separable case: \( W(x, y) = (1 + x)^{-\alpha}(1 + y)^{-\alpha}, \alpha \in \{1.5, 2, 3, 4, 5\} \).
(2) Non-separable case: \( W(x, y) = (1 + x + y)^{-\alpha}, \alpha \in \{2, 2.6, 3, 4, 5\} \).

Here the non-separable case with \( \alpha = 2 \) is included to make comparisons with the others so that we can examine how important the condition in Definition 1(a) is. For each combination of \( W(\cdot, \cdot) \) and \( \alpha \), we simulate 500 replications with \( t = 1000 \). From Veitch and Roy (2015); Borgs et al. (2018), we have that when \( t = 1000 \), the number of edges in the network is of order \( O(t^2) = O(10^6) \). Also, by [Caron et al. (2020), Theorem 4], the number of nodes in the graph is of order \( O(t^{1+1/\alpha}) \) in the separable case and \( O(t^{1+1/(\alpha-1)}) \) in the non-separable case. Hence, we think \( t = 1000 \) is sufficiently large to generate large graphs to test our asymptotic results.

Here all simulated replications have a similar distribution shape as depicted in Fig. 1. In other words, when plotted on a log-log scale, the empirical distribution shows a linearly decaying pattern as long as we stay away from the cut-off. Therefore, we estimate the power-law index by the absolute value of the slope estimate from a simple linear regression. To avoid the cut-off area, we choose a proportion of the empirical distribution such that the fitted regression returns an \( R^2 \geq 99.5\% \). This is done by first fitting a linear regression to the entire empirical distribution, then leaving out observations in the cut-off area until the target \( R^2 \) is achieved.

Following the method described above, we collect all tail index estimates (slope estimates) for the distribution of common connections, and plot them as boxplots in Fig. 2. The left panel contains all simulated separable examples, i.e. \( W(x, y) = (1 + x)^{-\alpha}(1 + y)^{-\alpha}, \alpha \in \{1.5, 2, 3, 4, 5\} \).
$\alpha \in \{1.5, 2, 3, 4, 5\}$, whereas all non-separable examples, i.e. $W(x, y) = (1 + x + y)^{-\alpha}$, $\alpha \in \{2, 2.6, 3, 4, 5\}$, are given in the right panel of Fig. 2. Both separable and non-separable cases reveal the same pattern that when $W(\cdot, \cdot) = (1 + x + y)^{-\alpha}$, $\alpha \in \{2, 2.6, 3, 4, 5\}$, are given in the right panel of Fig. 2. Both separable and non-separable cases reveal the same pattern that when $W(\cdot, \cdot)$ decays at a faster rate, i.e. when $\alpha$ is larger, the distribution of common connections will have a heavier tail. Similar pattern is also observed from the empirical degree distribution as given in Caron and Fox (2017); Caron et al. (2020); Veitch and Roy (2015).

To assess the tightness of bounds provided in the previous section, we summarize the mean, variance and range based on the 500 replications for each combination of $W(\cdot, \cdot)$ and $\alpha$. For the separable case, we compute the coverage proportion for the derived upper and lower bounds for the tail index:

$$\left[1 + \frac{1}{\alpha}, \max \left(\frac{3}{2} + \frac{1}{\alpha}, 1 + \frac{2}{\alpha}\right)\right],$$

(37)

i.e., the proportion of estimated tail indices that are within the interval in (37). Numerical results are given in Table 1. Overall, the mean and standard deviation are both decreasing as $\alpha$ increases, and the coverage proportion remains high, as long as $\alpha$ is neither too small nor too large. We also notice that as $\alpha$ increases, the estimated tail indices are gradually approaching the lower bound given in (37). Especially when $\alpha = 5$, the bound given by (37) is $[1.2, 1.7]$, so that 15.2% of the results fall within the small interval $[1.182, 1.2]$. This contributes to the slightly lower coverage proportion for the $\alpha = 5$ case. Having all these observations in mind, we conclude that the bounds indicated in (37) perform well in these examples, even though we now take all $(x, y) \in \Pi^0_\alpha$ into consideration.

For the non-separable case, we examine the tightness of the derived upper bound for the tail index, i.e.,

$$\left(1 + \frac{2}{(2\alpha - 1)}, 1 + \frac{4}{\alpha}\right),$$

(38)

in Table 2. Similar to the separable case, both mean and standard deviation are decreasing as $\alpha$ increases. Except the $\alpha = 2$ case, all other coverage proportions reach 100%. The relative poor coverage proportion for $\alpha = 2$ may be due to the fact that $\int_0^\infty \mu_1(x)dx = \infty$ when $\alpha = 2$.

Comparing Tables 1 and 2, we see that for the same $\alpha$, a separable $W(\cdot, \cdot)$ tends to generate a heavier tail for the common connection distribution than that in a non-separable $W(\cdot, \cdot)$ case. Note that although in both cases $W \in \mathcal{MRV}(-\alpha, \mathbb{R}^2_+ \setminus \{0\})$, these two models tend to generate structures which are quite different, for example, we obtain different behavior of graphex marginals $\mu_1$ and $\mu_2$ in the two cases. Using Theorem 1, for the non-separable $W$ we have $\mu_1 \in \mathcal{RV}_-(\alpha - 1)$. On the other hand, for the separable case, $W \in \mathcal{MRV}(-\alpha, \mathbb{R}_+^2 \setminus \{0\})$ and moreover $W \in \mathcal{MRV}(-2\alpha, (0, \infty)^2)$, and from Lemma 1 we have $\mu_1 \in \mathcal{RV}_{-\alpha}$. In order

| $\alpha$ | Mean  | Std Dev | Range             | Coverage (%) |
|---------|-------|---------|------------------|--------------|
| 1.5     | 2.095 | 0.052   | [1.980, 2.193]   | 88.8         |
| 2       | 1.769 | 0.056   | [1.689, 1.879]   | 100          |
| 3       | 1.438 | 0.039   | [1.366, 1.508]   | 100          |
| 4       | 1.322 | 0.041   | [1.249, 1.394]   | 99           |
| 5       | 1.238 | 0.033   | [1.182, 1.308]   | 84.8         |
to compare models with similar average connection behaviors, we tabulate in Table 3 the means and ranges of the estimated slope, where the tail indices are the same in the two models. Even here we observe that the common connection distribution tends to have heavier tails in the separable case. A similar comparison matching indices in the separable ($\mu_2 = 1.5$ with $\mu_2$ index $2\alpha = 3$) and non-separable ($\alpha = 2$ with $\mu_2$ index $2\alpha - 1 = 3$) cases also leads to observing a heavier tail behavior in the former.

### Table 3

| $\alpha$ | Mean | Std Dev | Range       | Coverage (%) |
|----------|------|---------|-------------|--------------|
| 2        | 2.228| 0.045   | [2.140, 2.310] | 81.4         |
| 2.6      | 2.121| 0.049   | [2.022, 2.210] | 100          |
| 3        | 2.081| 0.041   | [1.992, 2.150] | 100          |
| 4        | 1.878| 0.040   | [1.781, 1.950] | 100          |
| 5        | 1.754| 0.045   | [1.670, 1.829] | 100          |

### Table 2

| $\alpha$ | Mean | Std Dev | Range       | Coverage (%) |
|----------|------|---------|-------------|--------------|
| $\mu_1$  | Mean | Std Dev | Range       | Coverage (%) |
| 2        | 1.769|         | [1.689, 1.879] |              |
| 3        | 1.438|         | [1.366, 1.508] |              |
| 4        | 1.322|         | [1.249, 1.394] |              |

### 5 Conclusion

In this paper, we work under the paradigm of sparse exchangeable graphs generated by multivariate regularly varying graphex functions and investigate the distributional behavior of common connections between pairs of vertices. We conclude that the asymptotic distribution of number of common connections of randomly chosen pair of vertices, given that the vertices chosen are above an (increasing) threshold has a power-law behavior under certain regularity conditions. We distinguish between two types of generating graphex functions here, separable and non-separable, each of which leads to different asymptotic tail behavior. We manage to derive bounds for the tail behavior but not the exact rate, and our simulation studies verify the bounds obtained.

One future direction would be to find appropriate minimal conditions on the graphex function, so that we obtain tighter bounds for the tail rate of the distribution of number of common connections of randomly chosen pair of vertices. Network scientists have found this measure a good indicator for link-prediction between vertices. We believe understanding 2-c-degree functions, and possibly k-c-degree functions will also lead us towards a generating mechanism for networks like Facebook and LinkedIn, where connections are often made through friendship recommendation.
Acknowledgements The work of B. Das and G. Dai was supported by Ministry of Education Academic Research Fund Tier 2 grant MOE2017-T2-2-161 on “Learning from Common Connections in Social Networks”.

Data Availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

Aldous DJ (1981) Representations for partially exchangeable arrays of random variables. J Multivar Anal 11(4):581–598
Barabási A, Jeong H, Nédé Z, Ravasz E, Schubert A, Vicsek T (2002) Evolution of the social network of scientific collaborations. Physica A: Statistical mechanics and its applications 311(3):590–614
Bingham NH, Goldie CM, Teugels JL (1989) Regular variation, volume 27 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge
Bloznelis M, Kurauskas V (2016) Clustering function: another view on clustering coefficient. J Complex Networks 4(1):61–86
Borgs C, Chayes JT, Cohn H, Holden N (2018) Sparse exchangeable graphs and their limits via graphon processes. J Mach Learn Res 18(210):1–71
Borgs C, Chayes JT, Cohn H, Veitch V (2019) Sampling perspectives on sparse exchangeable graphs. Ann Probab 47(5):2754–2800
Caron F, Fox EB (2017) Sparse graphs using exchangeable random measures. J R Stat Soc Ser B Stat Methodol 79(5):1295–1366
Caron F, Panero F, Rousseau J (2020) On sparsity, power-law and clustering properties of graphex processes. Arxiv preprint arXiv:1708.03120
Das B, Ghosh S (2021) Growth of common friends in a preferential attachment model. Stoch Model 37(3):427–447
de Haan L, Resnick SI (1987) On regular variation of probability densities. Stochastic Processes and their Applications 25:83–93
Diaconis P, Janson S (2008) Graph limits and exchangeable random graphs. Rendiconti di Matematica e delle sue Applicazioni. Serie VII 28(1):33–61
Gradshteyn I, Ryzhik I (2007) Table of Integrals, series, and products. Academic Press, 7th edition
Gupta P, Goel A, Lin J, Sharma A, Wang D, Zadeh R (2013) WTF: The who to follow service at twitter. In: Proceedings of the 22Nd International Conference on World Wide Web, WWW’13. ACM, New York, NY, USA, pp 505–514
Herlau T, Schmidt MN, Mørup M (2016) Completely random measures for modelling block-structured sparse networks. In: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Eds) Advances in Neural Information Processing Systems, vol 29. Curran Associates, Inc., pp 4260–4268
Hoover DN (1979) Relations on probability spaces and arrays of random variables. preprint. https://www.stat.berkeley.edu/~aldous/Research/hoover.pdf
Janson S (2016) Graphons and cut metric on sigma-finite measure spaces. Arxiv preprint arXiv:1608.01833
Janson S (2017) On convergence for graphexes. Arxiv preprint arXiv:1702.06389
Kallenberg O (1990) Exchangeable random measures in the plane. J Theor Probab 3(1):81–136
Kingman JFC (1993) Poisson Processes. Oxford Studies in Probability. Clarendon Press, 1st ed
Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. J Am Soc Inf Sci Technol 58(7):1019–1031
Lovász L, Szegedy B (2006) Limits of dense graph sequences. Journal of Combinatorial Theory, Series B 96(6):933–957
Moller J, Waagepetersen RP (2003) Statistical inference and simulation for spatial point processes. CRC Press
Naulet Z, Roy DM, Sharma E, Veitch V (2021) Bootstrap estimators for the tail-index and for the count statistics of graphex processes. Electronic Journal of Statistics 15(1):282–325
Newman MEJ (2001) Clustering and preferential attachment in growing networks. Phys Rev E 64(2):025102
Resnick S (1987) Extreme Values, Regular Variation and Point Processes. Springer-Verlag, New York
Resnick S (2007) Heavy Tail Phenomena: Probabilistic and Statistical Modeling. Springer-Verlag, New York
Stam A (1977) Regular variation in \( \mathbb{R}^d \) and the Abel-Tauber theorem. Technical Report, unpublished, Mathematisch Instituut, Rijksuniversiteit Groningen, August 1977
Todeschini A, Miscouridou X, Caron F (2020) Exchangeable random measures for sparse and modular graphs with overlapping communities. J R Stat Soc Ser B (Stat Methodol) 82(2):487–520
Veitch V, Roy DM (2015) The class of random graphs arising from exchangeable random measures. Arxiv preprint arXiv:1512.03099
Veitch V, Roy DM (2019) Sampling and estimation for (sparse) exchangeable graphs. Ann Stat 47(6):3274–3299
Voitalov I, van der Hoorn P, van der Hofstad R, Krioukov D (2019) Scale-free networks well done. Physical Review Research 1(3):033034

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.