Probability in the Engineering and Informational Sciences (2025), **39**:3 370–396
doi:10.1017/S0269964824000160

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Graph connectivity with fixed endpoints in the random-connection model

Qingwei Liu and Nicolas Privault

Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371, Singapore
Corresponding author: Nicolas Privault; Email: nprivault@ntu.edu.sg and xiaozong30@gmail.com

Keywords: connectivity; cumulant method; normal approximation; Poisson point process; random graphs; random-connection model

MSC: 60D05; 05C80; 60G55; 60F05

Abstract
We consider the count of subgraphs with an arbitrary configuration of endpoints in the random-connection model based on a Poisson point process on $\mathbb{R}^d$. We present combinatorial expressions for the computation of the cumulants and moments of all orders of such subgraph counts, which allow us to estimate the growth of cumulants as the intensity of the underlying Poisson point process goes to infinity. As a consequence, we obtain a central limit theorem with explicit convergence rates under the Kolmogorov distance and connectivity bounds. Numerical examples are presented using a computer code in SageMath for the closed-form computation of cumulants of any order, for any type of connected subgraph, and for any configuration of endpoints in any dimension $d \geq 1$. In particular, graph connectivity estimates, Gram–Charlier expansions for density estimation, and correlation estimates for joint subgraph counting are obtained.

1. Introduction

This paper considers the statistics and asymptotic behavior of subgraph counts in a multidimensional random-connection model based on a Poisson point process, which can be used to model physical systems in, for example, statistical mechanics [9], wireless networks [8, 19, 28], or cosmology [5, 7].

The random-connection model, in which vertices are randomly located and connected with location-dependent probabilities, is a natural generalization of, for example, the Erdős–Rényi random graph or the stochastic block model [27]. Namely, given μ a diffuse Radon measure on $\mathbb{R}^d$, the random-connection model $G_H(\eta)$ consists of an underlying Poisson point process η on $\mathbb{R}^d$ with intensity of the form $\lambda\mu(\mathrm{d}x)$, $\lambda > 0$, in which any two vertices x, y in η are connected with the probability $H(x, y)$, where $H : \mathbb{R}^d \times \mathbb{R}^d \to [0, 1]$ is a symmetric connection function.

In addition to modeling the random locations of network nodes, many applications of wireless networks require the use of endpoints which are physical devices placed at given fixed locations, such as for example roadside units in vehicular networks such as VANETs, see, for example, [20] and [29]. The count of subgraphs that connect any single point x in the Poisson process η to m fixed endpoints $y_1, \ldots, y_m \in \mathbb{R}^d$ is known to have a Poisson distribution with mean $\lambda \int_{\mathbb{R}^d} H(x, y_1) \cdots H(x, y_m)\, \mu(\mathrm{d}x)$, see, for example, Section 4 in [23]. This Poisson property has been used in [13] to derive closed-form estimates of two-hop connectivity in the random-connection model when $m = 2$, see Proposition III.2 therein.

In this paper, we consider the count of general connected subgraphs with a general configuration of fixed endpoints at fixed locations $y_1, \ldots, y_m \in \mathbb{R}^d$ in the random-connection model $G_H(\eta \cup \{y_1, \ldots, y_m\})$

constructed on the union of the Poisson point process η and $\{y_1, \ldots, y_m\}$. In particular, we extend the subgraph count cumulant formulas obtained on $G_H(\eta)$ in [16] by taking into account the presence of endpoints in $G_H(\eta \cup \{y_1, \ldots, y_m\})$, and we provide SageMath coding implementations for joint cumulant expressions of any order.

In Proposition 3.4, we derive general expressions for the moments and cumulants of the count $N^G_{y_1,\ldots,y_m}$ of subgraphs with fixed endpoints $y_1, \ldots, y_m$ in $G_H(\eta \cup \{y_1, \ldots, y_m\})$. Such expressions allow us to determine the dominant terms in the growth of cumulants as the intensity λ of the underlying point process tends to infinity, by estimating the counts of vertices and edges in connected partition diagrams as in, for example, [14]. As a consequence, in Theorem 4.2, we obtain growth estimates for the cumulants of the subgraph count $N^G_{y_1,\ldots,y_m}$.

This allows us to show the convergence of renormalized subgraph counts to the normal distribution in Proposition 4.4 as the intensity λ of the underlying Poisson point process on $\mathbb{R}^d$ tends to infinity. Convergence rates under the Kolmogorov distance are then obtained in Proposition 4.5 for the normal approximation of subgraph counts from the combinatorics of cumulants and the Statulevičius condition, see [6, 25] and Lemma B.1, extending the results obtained in [16] for subgraphs without endpoints, see also [10] for other applications of this condition to concentration inequalities, normal approximation, and moderate deviations for random polytopes. In Proposition 4.6, connectivity probability estimates and bounds are derived using the second moment method and the factorial moment expansions in Proposition B.2.

In Section 5, we consider several examples of subgraphs with endpoints such as k-hop paths, triangles, and trees, for which exact cumulant computations are matched to their Monte Carlo estimates using the Rayleigh connection function $H(x, y) = e^{-\beta\|x-y\|^2}$, $\beta > 0$. In those examples, we obtain graph connectivity estimates, and correlation estimates for joint graph counting, which are matched to the outputs of Monte Carlo simulations. In addition, using third order cumulant expressions, we also provide improved fits of probability density functions of renormalized subgraph counts when the Gaussian approximation is not valid, see Figure 9.

Computations are done in closed form using symbolic calculus in the SageMath coding implementations presented in Appendices D–E, and available for download at https://github.com/nprivaul/random-connection. We note that although intensive computations may be required, the types of connected subgraphs and associated configurations of endpoints considered are only limited by the available computing power.

This paper is organized as follows. Section 2 introduces some preliminaries on subgraph counting and the computation of moments using summations over partitions in the random-connection model. In Section 3, we use partition diagrams to compute the cumulants of the counts of subgraphs with endpoints in the random-connection model. Subgraph count asymptotics and the associated central limit theorem are given in Section 4, and numerical examples are presented in Section 5. A general derivation of joint cumulant identities is given in Appendix A, extending the construction of [16] from the univariate to the multivariate case, for use in Section 5.5. Basic results on Gram–Charlier expansions and probability approximation using cumulant and moment methods are recalled in Appendices B and C. The SageMath codes for the computation of cumulants and joint cumulants are listed in Appendices D and E, and available for download at https://github.com/nprivaul/random-connection.

2. Subgraph counts in the random-connection model

In what follows we consider a Radon measure μ on $\mathbb{R}^d$, and we let $\mathbb{P}_\lambda$, $\lambda > 0$, denote the distribution of the Poisson point process η with intensity $\lambda\mu(\mathrm{d}x)$ on the space

$$\mathcal{C} := \left\{\eta \subset \mathbb{R}^d \ : \ |\eta \cap A| < \infty \text{ for any bounded set } A \subset \mathbb{R}^d\right\}$$

of locally finite configurations on $\mathbb{R}^d$, whose elements $\eta \in \Omega$ are identified with the Radon point measures $\eta = \sum_{x\in\eta} \epsilon_x$, so that $\eta(B)$ represents the random number of points contained in a Borel set in $\mathbb{R}^d$. In other words,

(1) for any relatively compact Borel set $B \subset \mathbb{R}^d$, the distribution of $\eta(B)$ under $\mathbb{P}_\lambda$ is Poisson with parameter $\lambda\mu(B)$;
(2) for any $n\geq 2$ and pairwise disjoint relatively compact Borel sets $B_1, \ldots, B_n \subset \mathbb{R}^d$, the random variables $\eta(B_1), \ldots, \eta(B_n)$ are independent under $\mathbb{P}_\lambda$.

For $n\geq 1$ we let $[n] := \{1, \ldots, n\}$, where n will later on denote the order of the considered moments and cumulants of subgraph counts, and for any set A we denote by $\Pi(A)$ the collection of all set partitions of A. We also let $|A|$ denote the number of elements of any finite set A, and, in particular, $|\sigma|$ represents the number of blocks in a partition $\sigma \in \Pi([n] \times [r])$. Our approach to the computation of moments relies on moment identities on the following form, see Proposition 3.1 in [22] and Proposition A.5 for its multivariate generalization.

Proposition 2.1. *Let $n \geq 1$ and $r \geq 1$, and let $f : (\mathbb{R}^d)^r \to \mathbb{R}$ be a sufficiently integrable measurable function. We have*

$$\mathbb{E}\left[\left(\sum_{(x_1,\ldots,x_r)\in\eta^r} f(x_1,\ldots,x_r)\right)^n\right] = \sum_{\rho\in\Pi([n]\times[r])} \lambda^{|\rho|} \int_{(\mathbb{R}^d)^{|\rho|}} \prod_{k=1}^n f\left(x_{\zeta_\rho(k,1)},\ldots,x_{\zeta_\rho(k,r)}\right) \mu(\mathrm{d}x_1)\cdots\mu(\mathrm{d}x_{|\rho|}),$$

where, for $\rho = \{\rho_1, \ldots, \rho_{|\rho|}\}$ a partition of $[n] \times [r]$, we let $\zeta_\rho(k, l)$ denote the index p of the block ρ_p of ρ to which (k, l) belongs.

In particular, Proposition 2.1 will yield cumulant expressions from Möbius inversion and combinatorial arguments based on [18], [14], and [16], see Propositions 3.5 and A.8.

Definition 2.2. *Given $H : \mathbb{R}^d \times \mathbb{R}^d \to [0, 1]$ a symmetric connection function and $y_1, \ldots, y_m$ fixed points in $\mathbb{R}^d$, the random-connection model $G_H(\eta \cup \{y_1, \ldots, y_m\})$ is the random graph built on the union of $\{y_1, \ldots, y_m\}$ and a Poisson point process sample η, in which any two distinct points $x, y \in \eta \cup \{y_1, \ldots, y_m\}$ are independently connected by an edge with the probability $H(x, y)$.*

In the sequel, we will consider a family of connected graphs with endpoints which are described in the following assumption.

Assume 2.3. *Given $r\geq 2$ and $m\geq 0$, we consider a connected graph $G = (V_G, E_G)$ with edge set E_G and vertex set $V_G = (v_1, \ldots, v_r; w_1, \ldots, w_m)$, such that*

(i) the subgraph G induced by G on $\{v_1, \ldots, v_r\}$ is connected, and
(ii) the endpoint vertices $w_1, \ldots, w_m$ are not adjacent to each other in G.

In case $m = 0$, Condition (ii) is void and $V_G = (v_1, \ldots, v_r)$.

In Figure 1, an example of a graph satisfying Assume 2.3 is described with $r = 4$ and $m = 2$. As a convention, in the next definition the sets $\{w_1, \ldots, w_m\}$ and $\{y_1, \ldots, y_m\} \subset \mathbb{R}^d$ are empty when $m = 0$.

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

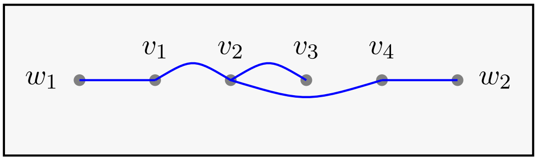

Figure 1. *Graph* $G = (V_G, E_G)$ *with* $V_G = (v_1, v_2, v_3, v_4; w_1, w_2)$, n = *3,* r = *4,* m = *2.*

Definition 2.4. *Let G be a graph satisfying Assumption 2.3. Given* $m \geq 0$ *fixed points* $y_1, \ldots, y_m \in \mathbb{R}^d$*, for almost surely* η *we let* $N^G_{y_1,\ldots,y_m}$ *denote the count of subgraphs in* $G_H(\eta \cup \{y_1, \ldots, y_m\})$ *that are isomorphic to* $G = (V_G, E_G)$ *in the sense that there exists a (random) injection from* V_G *into* $\eta \cup \{y_1, \ldots, y_m\}$ *which is one-to-one from* $\{w_1, \ldots, w_m\}$ *to* $\{y_1, \ldots, y_m\}$*, and preserves the graph structure of G.*

According to Definition 2.4, we express the subgraph count $N^G_{y_1,\ldots,y_m}$ as

$$N^G_{y_1,\ldots,y_m} = \sum_{(x_1,\ldots,x_r)\in\eta^r} f_{y_1,\ldots,y_m}(x_1, \ldots, x_r),$$

where the random function $f : (\mathbb{R}^d)^r \to \{0, 1\}$ defined as

$$f_{y_1,\ldots,y_m}(x_1, \ldots, x_r) := \prod_{\substack{1\leq i\leq r\\ 1\leq j\leq m\\ \{v_i,w_j\}\in E_G}} \mathbf{1}_{\{y_j\leftrightarrow x_i\}} \prod_{\substack{1\leq k,l\leq r\\ \{v_k,v_l\}\in E_G}} \mathbf{1}_{\{x_k\leftrightarrow x_l\}}, \qquad x_1, \ldots, x_r \in \mathbb{R}^d,$$

is independent of the Poisson point process η, and $\mathbf{1}_{\{x\leftrightarrow y\}} = 1$ if and only if $x \neq y$ and $x, y \in \mathbb{R}^d$ are connected in the random-connection model $G_H(\eta \cup \{y_1, \ldots, y_m\})$.

3. Partition diagrams

This section introduces the combinatorial background needed for the derivation of moment and cumulant expressions of subgraph counts. The next definition introduces a notion of connectedness over the rows of partitions of $[n] \times [r]$, and a flatness property which is satisfied when two indices on a same row belong to a given block, see Chapter 4 of [21] and Figure 2.

Definition 3.1. *Given* $n, r \geq 1$*, let* $\pi := \{\pi_1, \ldots, \pi_n\}$ *be the partition in* $\Pi([n] \times [r])$ *given by*

$$\pi_i := \{(i, 1), \ldots, (i, r)\}, \quad i = 1, \ldots, n.$$

(i) A set partition $\sigma \in \Pi([n] \times [r])$ *is connected if* $\sigma \vee \pi = \widehat{1}$*, where* $\sigma \vee \pi$ *is the finest set partition which is coarser than both* σ *and* π*, and* $\widehat{1} = \{[n] \times [r]\}$ *is the coarsest partition of* $[n] \times [r]$*.*
(ii) A set partition $\sigma \in \Pi([n] \times [r])$ *is non-flat if* $\sigma \wedge \pi = \widehat{0}$*, where* $\sigma \wedge \pi$ *is the coarsest set partition which is finer than both* σ *and* π*, and* $\widehat{0}$ *is the finest partition of* $[n] \times [r]$*.*

We let $\Pi_{\widehat{1}}([n] \times [r])$ *denote the collection of all connected partitions of* $[n] \times [r]$*.*

In the sequel, every partition $\rho \in \Pi(\pi_1 \cup \cdots \cup \pi_n)$ will be arranged into a diagram denoted by $\Gamma(\rho, \pi)$, by arranging $\pi_1, \ldots, \pi_n$ into n rows and connecting together the elements of every block of ρ. Figure 2 presents two illustrations of flat non-connected and connected non-flat partition diagrams with $n = 5$ and $r = 4$, in which the partition ρ is represented using line segments.

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

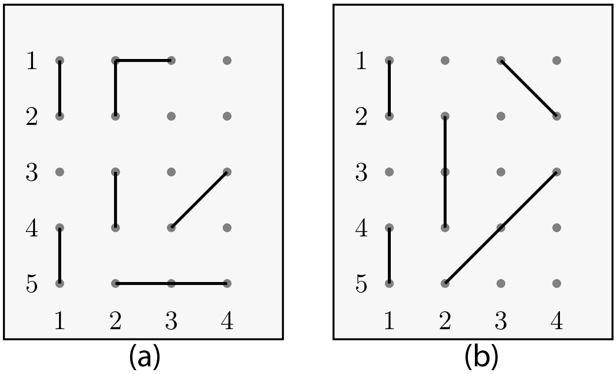

Figure 2. *Two examples of partition diagrams with* n = 5 *and* r = 4*. (a) Flat non-connected diagram* $\Gamma(\rho, \pi)$*. (b) Connected non-flat diagram* $\Gamma(\rho, \pi)$*.*

In Definition 3.2, to any graph G and set partition $\rho \in \Pi([n] \times [r])$, we associate a graph ρ_G whose vertices are the blocks of ρ. For this, we use n copies of the graphs induced by the v_i's with addition of the end-points $w_1, \ldots, w_m$, and we merge the nodes obtained in this way on $[n] \times [r]$ according to the partition ρ.

Definition 3.2. *Given* ρ *a partition of* $[n] \times [r]$ *and* $G = (V_G, E_G)$ *a connected graph on* $V_G = (v_1, \ldots, v_r; w_1, \ldots, w_m)$*, we let* ρ_G *denote the graph constructed as follows on* $[m] \cup [n] \times [r]$*:*

(i) for all $j_1, j_2 \in [r]$*,* $j_1 \neq j_2$*, and* $i \in [n]$*, an edge links* (i, j_1) *to* (i, j_2) *iff* $\{v_{j_1}, v_{j_2}\} \in E_G$*;*
(ii) for all $(j, k) \in [r] \times [m]$ *and* $i \in [n]$*, an edge links* (k) *to* (i, j) *iff* $\{v_j, w_k\} \in E_G$*;*
(iii) for all $i_1, i_2 \in [n]$ *and* $j_1, j_2 \in [r]$*, merge any two nodes* (i_1, j_1) *and* (i_2, j_2) *if they belong to a same block in* ρ*;*
(iv) eliminating any redundant edges created by the above construction.

If $\rho \in \Pi([n] \times [r])$ takes the form $\rho = \{b_1, \ldots, b_{|\rho|}\}$, the graph ρ_G forms a connected graph with $|\rho| + m$ vertices, and we reindex the set of vertices V_{ρ_G} of ρ_G as $V_{\rho_G} = [|\rho| + m]$ according to the lexicographic order on $\mathbb{N} \times \mathbb{N}$, followed by the remaining m vertices, indexed as $\{|\rho| + 1, \ldots, |\rho| + m\}$, see Figure 3(b) in which we have $|\rho| = 9$, $m = 2$, and $V_{\rho_G} = (1, \ldots, 9; 10, 11)$.

Example. Take $r = 4$, $m = 2$, and $V_G = (v_1, v_2, v_3, v_4; w_1, w_2)$. Figure 3(b) shows the graph ρ_G defined from $G = (V_G, E_G)$ of Figure 1 and the 9-block partition $\rho \in \Pi([3] \times [4])$ given by

$$\begin{aligned}\rho = \big\{&\{(1,1)\},\\ &\{(1,2),(2,2)\},\\ &\{(1,3)\},\\ &\{(1,4)\},\\ &\{(2,1),(3,1)\},\\ &\{(2,3)\},\\ &\{(2,4),(3,4)\},\\ &\{(3,2)\},\\ &\{(3,3)\}\big\}.\end{aligned}$$

In Figure 3(a), the partition ρ is represented using line segments.

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

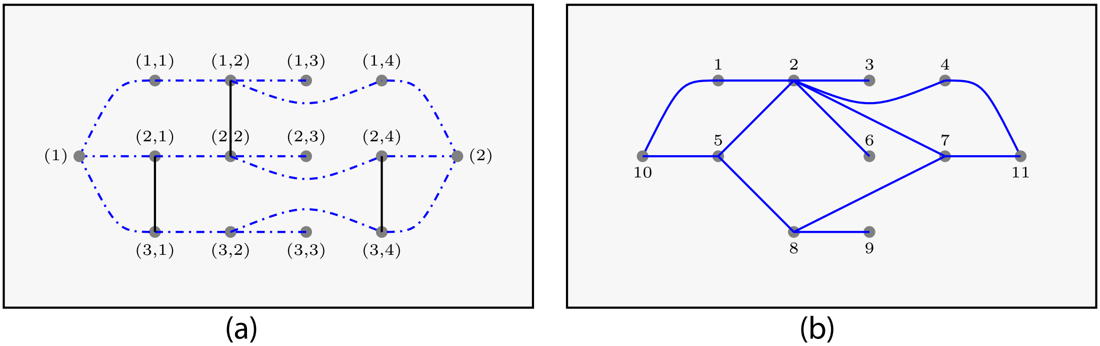

Figure 3. *Example of graph ρ_G with $n = 3$, $r = 4$, and $m = 2$. (a) Diagram before merging edges and vertices. (b) Graph ρ_g after merging edges and vertices.*

Definition 3.3. *For $\rho \in \Pi([n] \times [r])$ of the form $\rho = \{b_1, \dots, b_{|\rho|}\}$ and $j \in [m]$, we let*

$$\mathcal{A}_j^\rho := \{k \in [|\rho|] \ : \ \exists (s,i) \in b_k \text{s.t.} (v_i, w_j) \in E_G\}$$

denote the neighborhood of the vertex $(|\rho| + j)$ in ρ_G, $j = 1, \dots, m$.

For example, in the graph ρ_G of Figure 3 we have $\mathcal{A}_1^\rho = \{1, 5\}$ and $\mathcal{A}_2^\rho = \{4, 7\}$. The following partition summation formulas extend [16, Prop. 5.1] to the counting of subgraphs with endpoints, and they are a special case of Proposition A.8 in appendix, which deals with joint subgraph counting.

Proposition 3.4. *Let $m \geq 0$. The moments and cumulants of $N^G_{y_1,\dots,y_m}$ admit the following expressions:*

$$\mathbb{E}_\lambda\left[\left(N^G_{y_1,\dots,y_m}\right)^n\right] = \sum_{\substack{\rho \in \Pi([n]\times[r]) \\ \rho \wedge \pi = \widehat{0} \\ \text{(non-flat)}}} \lambda^{|\rho|} \int_{(\mathbb{R}^d)^{|\rho|}} \prod_{\substack{1 \leq j \leq m \\ i \in \mathcal{A}_j^\rho}} H(x_i, y_j) \prod_{\substack{1 \leq k,l \leq |\rho| \\ \{k,l\} \in E_{\rho_G}}} H(x_k, x_l)\, \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_{|\rho|}),$$

and

$$\kappa_n\left(N^G_{y_1,\dots,y_m}\right) = \sum_{\substack{\rho \in \Pi_{\widehat{1}}([n]\times[r]) \\ \rho \wedge \pi = \widehat{0} \\ \text{(non-flat connected)}}} \lambda^{|\rho|} \int_{(\mathbb{R}^d)^{|\rho|}} \prod_{\substack{1 \leq j \leq m \\ i \in \mathcal{A}_j^\rho}} H(x_i, y_j) \prod_{\substack{1 \leq k,l \leq |\rho| \\ \{k,l\} \in E_{\rho_G}}} H(x_k, x_l)\, \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_{|\rho|}). \quad (3.1)$$

We note in particular that $N^G_{y_1,\dots,y_m}$ has positive cumulants, and when $n = 1$ the first moment of $N^G_{y_1,\dots,y_m}$ is given by

$$\mathbb{E}_\lambda\left[N^G_{y_1,\dots,y_m}\right] = \lambda^r \int_{(\mathbb{R}^d)^r} \prod_{\substack{1 \leq i \leq r \\ 1 \leq j \leq m \\ \{v_i, w_j\} \in E_G}} H(x_i, y_j) \prod_{\substack{1 \leq k,l \leq r \\ \{v_k, v_l\} \in E_G}} H(x_k, x_l) \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_r).$$

The cumulant formula of Proposition 3.4 is implemented in the code listed in Appendix D. We also recall the following lemma, see [16, Lem. 2.8], in which maximality of connected non-flat partitions refers to maximizing the number of blocks.

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

Lemma 3.5.

(i) *The cardinality of the set* $\mathcal{C}(n,r)$ *of connected non-flat partitions of* $[n]\times[r]$ *satisfies*

$$|\mathcal{C}(n,r)| \leq n!^r r!^{n-1}, \qquad n,r\geq 1. \tag{3.2}$$

(ii) *The cardinality of the set* $\mathcal{M}(n,r)$ *of maximal connected non-flat partition of* $[n]\times[r]$ *satisfies*

$$|\mathcal{M}(n,r)| = r^{n-1}\prod_{i=1}^{n-1}(1+(r-1)i), \qquad n,r\geq 1,$$

with the bounds

$$((r-1)r)^{n-1}(n-1)! \leq |\mathcal{M}(n,r)| \leq ((r-1)r)^{n-1}n!, \quad n\geq 1,\ r\geq 2. \tag{3.3}$$

4. Subgraph count asymptotics

In this section, we let $m\geq 1$ and investigate the asymptotic behavior of the cumulants $\kappa_n\big(N^G_{y_1,\ldots,y_m}\big)$ in (3.1) as the intensity λ tends to infinity, which extends the treatment of [16] from $m=0$ to $m\geq 1$.

Assume 4.1. *We assume that*

(i) μ *is the Lebesgue measure on* $\mathbb{R}^d$*, and*
(ii) *the connection function* $H:\mathbb{R}^d\times\mathbb{R}^d\to[0,1]$ *is translation invariant, that is* $H(x,y)=H(0,y-x)$, $x,y\in\mathbb{R}^d$*, and*

$$\int_{\mathbb{R}^d} H(0,y)\mathrm{d}y < \infty.$$

The following result provides growth estimates for the cumulants of $N^G_{y_1,\ldots,y_m}$.

Theorem 4.2 *Let* $m\geq 1$, $n\geq 1$ *and* $r\geq 2$*, and suppose that Assume 4.1 is satisfied. We have*

$$0 < \kappa_n\big(N^G_{y_1,\ldots,y_m}\big) \leq n!^r r!^{n-1}(C\lambda)^{1+(r-1)n}, \tag{4.1}$$

and, for $n=2$*,*

$$(r-1)rc^{2r}\lambda^{2r-1} \leq \kappa_2\big(N^G_{y_1,\ldots,y_m}\big) \leq r!(C\lambda)^{2r-1}, \tag{4.2}$$

where $c,C>0$ *are constants independent of* $r\geq 2$ *and* $n\geq 2$.

Proof. According to Proposition 3.4, every non-flat connected partition $\rho\in\Pi([n]\times[r])$ corresponds to a summand of order $O(\lambda^{|\rho|})$. As the cardinality of maximal non-flat connected partitions is $1+(r-1)n$, the dominating asymptotic order is $O(\lambda^{1+(r-1)n})$. Precisely, by (3.2)–(3.3) and (3.1), letting

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

$j_0 \in \{1,\dots,m\}$ such that $\mathcal{A}^{\rho}_{j_0} \neq \emptyset$, for some $i_0 \in \mathcal{A}^{\rho}_{j_0}$ we have

$$\begin{aligned}
&c^{n|E_G|}C^{1+(r-1)n}((r-1)r)^{n-1}(n-1)!\lambda^{1+(r-1)n} \\
&\quad \leq \kappa_n\big(N^G_{y_1,\dots,y_m}\big) \\
&\quad \leq \lambda^{1+(r-1)n} \sum_{\substack{\rho\in\Pi_{\widehat{1}}([n]\times[r]) \\ \rho\wedge\pi=\widehat{0} \\ \text{(non-flat connected)}}} \int_{(\mathbb{R}^d)^{|\rho|}} \prod_{\substack{1\leq j\leq m \\ i\in\mathcal{A}^{\rho}_j}} H(x_i,y_j) \prod_{\substack{1\leq k,l\leq|\rho| \\ \{k,l\}\in E_{\rho_G}}} H(x_k,x_l)\mathrm{d}x_1\cdots\mathrm{d}x_{|\rho|} \\
&\quad \leq \lambda^{1+(r-1)n} \sum_{\substack{\rho\in\Pi_{\widehat{1}}([n]\times[r]) \\ \rho\wedge\pi=\widehat{0} \\ \text{(non-flat connected)}}} \int_{(\mathbb{R}^d)^{|\rho|}} H(x_{i_0},y_{j_0}) \prod_{\substack{1\leq k,l\leq|\rho| \\ \{k,l\}\in E_{\rho_G}}} H(x_k,x_l)\mathrm{d}x_1\cdots\mathrm{d}x_{|\rho|} \\
&\quad \leq \lambda^{1+(r-1)n} \sum_{\substack{\rho\in\Pi_{\widehat{1}}([n]\times[r]) \\ \rho\wedge\pi=\widehat{0} \\ \text{(non-flat connected)}}} \int_{(\mathbb{R}^d)^{|\rho|}} H(x_{i_0},y_{j_0}) \prod_{\substack{1\leq k,l\leq|\rho| \\ \{k,l\}\in E_{\rho'_G}}} H(x_k,x_l)\mathrm{d}x_1\cdots\mathrm{d}x_{|\rho|},
\end{aligned}$$

where for every $\rho \in \Pi_{\widehat{1}}([n]\times[r])$, ρ'_G is a spanning tree contained in ρ_G, with vertices $\{1,\dots,|\rho|,|\rho|+j\}$ and such that $|\rho|+j_0$ is a leaf. By integrating successively on the variables which correspond to leaves of ρ'_G as in the proofs of, for example, Theorem 7.1 of [15] or Lemma 3.1 of [3] and using (3.2), we obtain

$$\kappa_n\big(N^G_{y_1,\dots,y_m}\big) \leq (C\lambda)^{1+(r-1)n} n!^r r!^{n-1},$$

due to Assume 4.1-(ii), where $C := \max\big(1, \int_{\mathbb{R}^d} H(0,y)dy\big)$, which yields the right-hand side (4.1). In addition, Proposition 3.4 shows that all cumulants are positive, which completes the proof of (4.1). On the other hand, when $r\geq 2$, by (3.3) we have

$$\kappa_2\big(N^G_{y_1,\dots,y_m}\big) \geq (r-1)rC^{2r}\lambda^{2r-1},$$

where $C>0$ is a constant independent of $r\geq 2$ and $n\geq 2$, which shows (4.2). □

In what follows, we consider the centered and normalized subgraph count cumulants defined as

$$\widetilde{N}^G_{y_1,\dots,y_m} := \frac{N^G_{y_1,\dots,y_m} - \kappa_1\big(N^G_{y_1,\dots,y_m}\big)}{\sqrt{\kappa_2\big(N^G_{y_1,\dots,y_m}\big)}}.$$

Corollary 4.3. *Let $m\geq 1$, $n\geq 2$ and $r\geq 2$. We have*

$$\left|\kappa_n\big(\widetilde{N}^G_{y_1,\dots,y_m}\big)\right| \leq n!^r C_r^{n/2}\lambda^{-(n/2-1)},$$

where $C_r > 0$ is a constant depending only on $r\geq 2$.

As a consequence of Corollary 4.3, the skewness of $\widetilde{N}^G_{y_1,\dots,y_m}$ satisfies

$$\left|\kappa_3\big(\widetilde{N}^G_{y_1,\dots,y_m}\big)\right| \leq C_r\lambda^{-1/2}, \tag{4.3}$$

where $C_r > 0$ is a constant depending only on $r\geq 2$.

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

Proposition 4.4. *Let $m \geq 1$. The renormalized subgraph count $\widetilde{N}^G_{y_1,\ldots,y_m}$ converges in distribution to the standard normal distribution $\mathcal{N}(0,1)$ as λ tends to infinity.*

Proof. From Corollary 4.3 and (4.3), we have $\kappa_1(\widetilde{N}^G_{y_1,\ldots,y_m}) = 0$, $\kappa_2(\widetilde{N}^G_{y_1,\ldots,y_m}) = 1$, and

$$\lim_{n\to\infty} \kappa_n(\widetilde{N}^G_{y_1,\ldots,y_m}) = 0, \qquad n \geq 3,$$

hence the conclusion follows from Theorem 1 in [11]. □

In addition, from Corollary 4.3 and Lemma B.1, the convergence result of Proposition 4.4 can be made more precise via the following convergence bound in the Kolmogorov distance, which extends Corollary 7.1 in [16] from $m=0$ to $m \geq 1$.

Proposition 4.5. *Let $m \geq 1$. We have*

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}_\lambda(\widetilde{N}^G_{y_1,\ldots,y_m} \leq x) - \Phi(x) \right| \leq C_r \lambda^{-1/(4r-2)}, \qquad r \geq 2,$$

where $C_r > 0$ is a constant depending only on $r \geq 2$ and Φ is the cumulative distribution function of the standard normal distribution.

By the second moment method, see, for example, (3.4) page 54 of [12] or Theorem 2.3 in [24], we also obtain the following lower bound for endpoint connectivity and subgraph existence.

Proposition 4.6. *Let $m \geq 1$. We have*

$$\mathbb{P}_\lambda(N^G_{y_1,\ldots,y_m} > 0) \geq \frac{\left(\mathbb{E}_\lambda[N^G_{y_1,\ldots,y_m}]\right)^2}{\mathbb{E}_\lambda\left[(N^G_{y_1,\ldots,y_m})^2\right]}, \qquad \lambda > 0. \tag{4.4}$$

Theorem 4.2 also shows the bounds

$$\frac{C_{r,1}}{\lambda} \leq \frac{\kappa_2(N^G_{y_1,\ldots,y_m})}{\left(\mathbb{E}_\lambda[N^G_{y_1,\ldots,y_m}]\right)^2} \leq \frac{C_{r,2}}{\lambda}, \qquad \lambda > 0, \tag{4.5}$$

for some constants $C_{r,1}, C_{r,2} > 0$ depending only on $r \geq 2$, from which it follows that the lower bound (4.4) converges to 1 as λ tends to infinity.

5. Numerical examples

In this section, we assume that H is the Rayleigh connection function

$$H_\beta(x,y) := e^{-\beta\|x-y\|^2}, \qquad x, y \in \mathbb{R}^d,$$

where $\beta > 0$, and μ is the Lebesgue measure on $\mathbb{R}^d$. In this case, Assume 4.1 is satisfied. In the following examples, the SageMath code listed in Appendix D is run after loading the definitions of Table 1. Computations in this and the following examples are run on a standard desktop computer with an 8-core CPU at 4.10 GHz. The limitations imposed by this hardware configuration constrain the product $n \times r \times d$ to be below 15 approximately, in order to maintain computation times at a reasonable level. The illustrations of Figures 4, 6, 8, and 10 are provided in dimension $d = 2$ for ease of visualization only.

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

Table 1. *Functions definitions.*

`load("cumulants_parallel.sage")`	# Loading the functions definitions
`λ,β= var("λ,β"); assume(β> 0)`	# Variable definitions
`def H(x,y,β): return exp(-β*(x-y)**2)`	# Connection function $H(x, y)$
`def mu(x,λ,β): return 1`	# Flat intensity of $\mu(dx)$

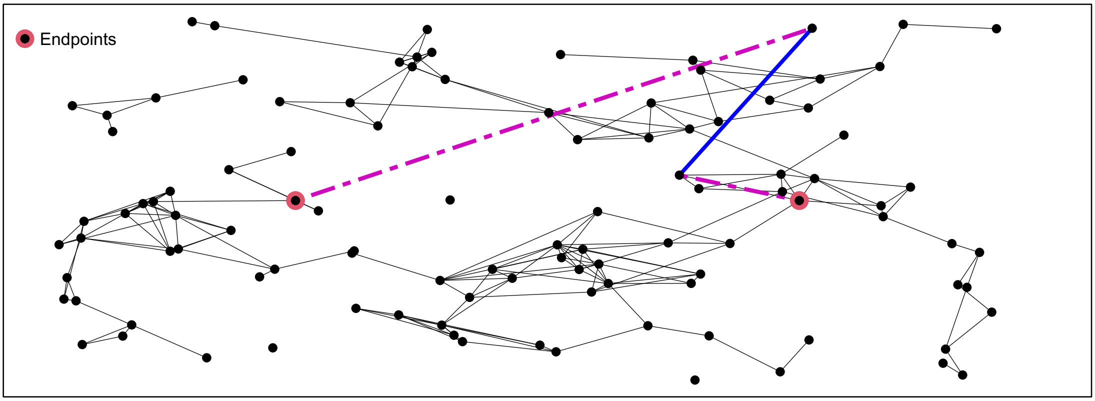

Figure 4. *A three-hop path with two endpoints in dimension* $\mathrm{d} = 2$*.*

Actual computations may be provided in lower dimension d, due to hardware performance constraints when the cumulant order n is beyond 4. Computations in dimension $d = 2$ are presented in Section 5.3 for triangles with endpoints and in Section 5.4 for trees with one endpoint cumulants of orders 2 and 3, while in Section 5.1 for three-hop paths with two endpoint and in Section 5.2 for four-hop paths with two endpoints we take $d = 1$ in order to reach the cumulant orders $n = 6$ and $n = 4$. In subsequent code inputs, graphs are coded by their edge set E_G, and the set of endpoints is given by the sequence $\mathsf{EP} = [\mathsf{EP}_1, \ldots, \mathsf{EP}_m]$, where EP_i denotes the set of vertices of the subgraph G on $\{v_1, \ldots, v_r\}$ which are attached to the ith endpoint, $i = 1, \ldots, m$, with $\mathsf{EP} := [\ ]$ the empty sequence when G has no endpoint $(m = 0)$.

5.1. Three-hop paths with two endpoints

By a k-hop path, we mean a non-self intersecting path having k edges. We take $m = 2$, $r = 2$, and in Table 2 we compute the first three cumulants of $N^G_{y_1,y_2}$ when G is a three-hop path with two endpoints in dimension $d = 1$, see Figure 4 for an illustration in dimension $d = 2$. Unlike in the two-hop with two endpoints case, this three-hop count does not have a Poisson distribution. In the following Figures 4, 6, 8, and 10, the endpoints are denoted by red dots, and their edges are denoted by purple dashed lines.
To make cumulant expressions more compact, the exact formulas in Table 2 are expressed with $y_1 = y_2 = 0$ and $\beta := \pi$, in dimension $d = 1$.
Table 3 lists the counts of connected non-flat partitions and runtimes for the computation of cumulants of orders 1–6 and shows that such partitions represent only a fraction (around 25%) of total partition counts.
Figure 5 presents connectivity estimates based on the moment and cumulant formulas of Propositions 4.6 and B.2, in dimension $d = 1$.

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

Table 2. *Cumulants of the count of two-hop paths with two endpoints in dimension* d = *1.*

G = [[1,2]]; EP =[[1],[2]]; d=1 # Single edge graph $r = 2$, two endpoint $m = 2$, dimension $d = 1$

Command	Order	Cumulant output	Connected non-flat partitions
c(1,d,G,EP,mu,H)	1st	$\frac{1}{\sqrt{3}}\lambda^2$	1
c(2,d,G,EP,mu,H)	2nd	$\left(\frac{1}{\sqrt{3}}+\frac{1}{\sqrt{2}}\right)\lambda^3+\left(\frac{1}{\sqrt{3}}+\frac{1}{2\sqrt{2}}\right)\lambda^2$	6
c(3,d,G,EP,mu,H)	3rd	$\left(\sqrt{\frac{12}{7}}+\frac{3}{\sqrt{5}}+\frac{3}{\sqrt{7}}+\frac{12}{\sqrt{31}}\right)\lambda^4+\left(\sqrt{3}+\sqrt{\frac{3}{2}}+\frac{17}{5\sqrt{2}}+\frac{12}{\sqrt{19}}\right)\lambda^3 + \left(\frac{3}{2\sqrt{2}}+\frac{1}{\sqrt{3}}\right)\lambda^2$	68

Table 3. *Computation times and counts of connected non-flat* vs. *all partitions in* $\Pi([n]\times[2])$.

Order n	2 blocks	3 blocks	4 blocks	5 blocks	6 blocks	7 blocks	Total	$\Pi([n]\times[2])$	Comp. time
1st	1	0	0	0	0	0	1	2	0.5 s
2nd	2	4	0	0	0	0	6	15	1 s
3rd	4	32	32	0	0	0	68	203	3 s
4th	8	208	624	352	0	0	1,192	4,140	1 m
5th	16	1,280	8,960	13,904	5,040	0	29,200	115,975	47 m
6th	32	7,744	116,160	375,776	351,456	88,544	939,712	4,213,597	29 hours

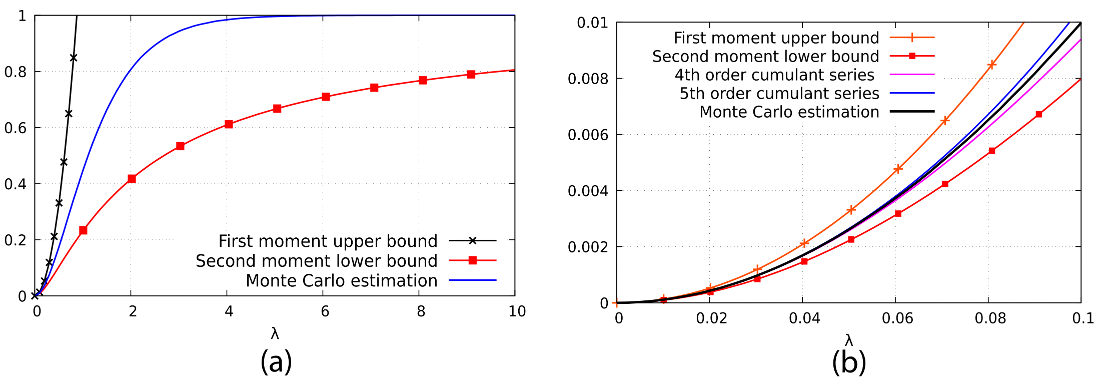

Figure 5. *Connection probabilities. (a) First and second moment bounds (4.4). (b) Cumulant approximations (B.1) with n = 0.*

5.2. Four-hop paths with two endpoints

Here, we take $m = 2$ and $r = 3$, and in Table 4 we compute the first cumulant of $N^G_{y_1,y_2}$ when G is a four-hop path with two endpoints in dimension $d = 1$, see Figure 6 for an illustration in dimension $d = 2$. The closed-form expressions in Table 4 are expressed with $y_1 = y_2 = 0$ and $\beta := \pi$, in dimension $d = 1$. Table 5 presents the counts of connected non-flat partitions at different orders, and shows that such partitions represent only a fraction (around 10%) of total partition counts.
In Figure 7, we plot the corresponding moment expressions *vs.* their Monte Carlo estimates in dimension $d = 1$, with the parameters of Table 4.

5.3. Triangles with endpoints

Taking $r = 3$ and $m = 3$, in Table 6, we compute the first and second cumulants of $N^G_{y_1,y_1,y_3}$ when G is a triangle with three endpoints in dimension $d = 1$, see Figure 8 for an illustration in dimension $d = 2$.

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

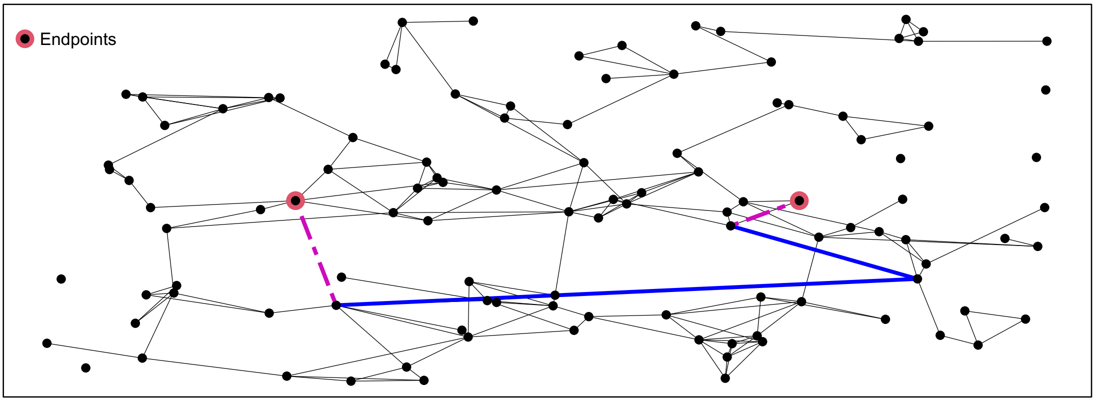

Figure 6. *A four-hop path with two endpoints in dimension* d = *2.*

Table 4. *First and second cumulants of the count of four-hop paths with two endpoints.*

G = [[1,2],[2,3]]; EP=[[1],[3]]; d=1		# four-hop path $r = 3$, two endpoints $m = 2$, dimension $d = 1$	
Instruction	Order	Cumulant output	Connected non-flat partitions
c(1,d,G,EP,mu,H)	1st	$\frac{\lambda^3}{2}$	1
c(2,d,G,EP,mu,H)	2nd	$\frac{1}{6}\lambda^3\left(\left(\sqrt{6}+4\sqrt{\frac{3}{5}}+\frac{3}{2\sqrt{2}}+\frac{12}{\sqrt{7}}\right)\lambda^2 + \left(3\sqrt{3}+16\sqrt{\frac{3}{7}}+8\sqrt{\frac{3}{11}}+\frac{3}{2\sqrt{2}}+\frac{6}{\sqrt{5}}\right)\lambda + \sqrt{3}+\sqrt{6}+6\right)$	33

Table 5. *Computation times and counts of connected non-flat* vs. *all partitions in* $\Pi([n]\times[3])$.

Order n	3 blocks	4 blocks	5 blocks	6 blocks	7 blocks	8 blocks	9 blocks	Total	$\Pi([n]\times[3])$	Comp. time
1st	1	0	0	0	0	0	0	1	5	1 s
2nd	6	18	9	0	0	0	0	33	203	2 s
3rd	36	540	1,242	864	189	0	0	2,871	21,147	4 m
4th	216	13,608	94,284	186,624	145,908	48,276	5,589	494,500	4,213,597	19 hours

The closed-form expressions in Table 6 are expressed with $y_1 = y_2 = y_3 = 0$ and $\beta := \pi$, in dimension $d = 1$.

Table 7 presents computation times in dimension $d = 2$.

Figure 9 presents second and third order Gram–Charlier expansions (C.1)–(C.2) for the probability density function of the count $N^G_{y_1,y_1,y_3}$ of triangles with three endpoints, based on exact second and third cumulant expressions.

In Figure 9, the purple areas correspond to probability density estimates obtained by Monte Carlo simulations with $\beta = 2$. The second order expansions correspond to the Gaussian diffusion approximation obtained by matching first and second order moments. Figure 9 shows that the actual probability density estimates obtained by simulation can be significantly different from their Gaussian diffusion approximations when skewness takes large absolute values. In addition, in Figure 9, the fourth order Gram–Charlier expansions appear to give the best fit to the actual probability densities, which have positive skewness.

5.4. Trees with one endpoint

Here we take $r = 4$ and $m = 1$, and in Table 8 we compute the first and second cumulants of $N^G_{y_1,y_1,y_3}$ when G is made of a tree and a single endpoint in dimension $d = 2$, see Figure 10 for an illustration.

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

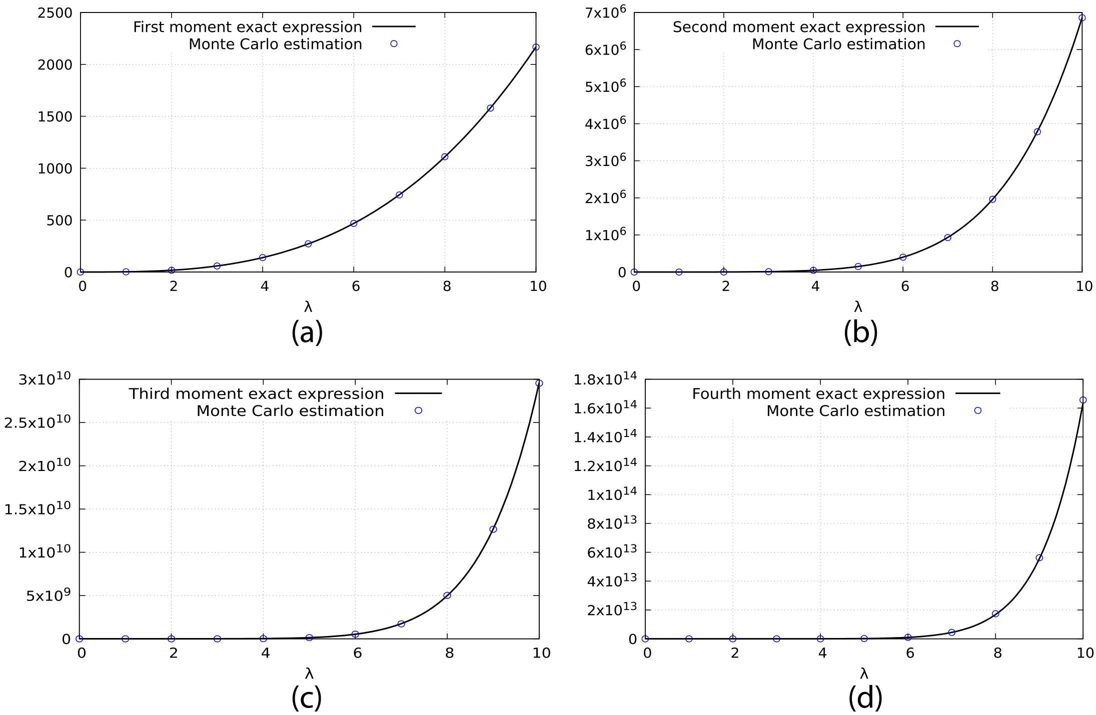

Figure 7. *Moment estimates. (a) First moment. (b) Second moment. (c) Third moment. (d) Fourth moment.*

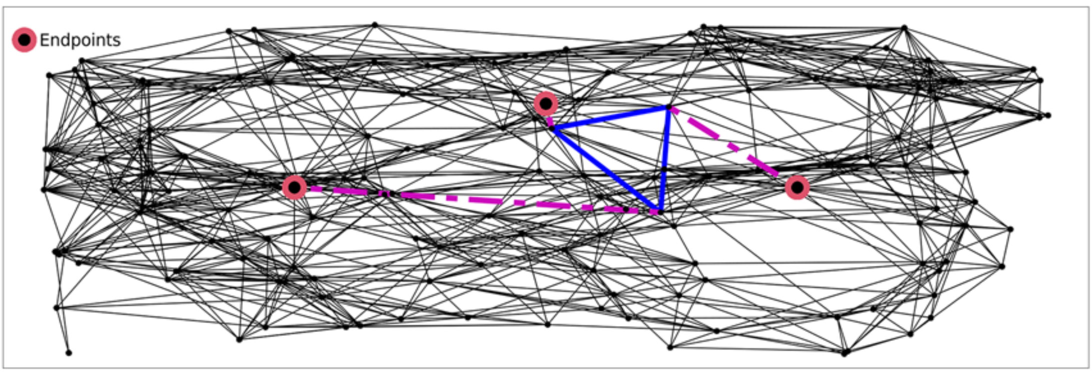

Figure 8. *A triangle with three endpoints in dimension* d = 2.

Table 6. *First and second cumulants of the count of triangles with three endpoints.*

G = [[1,2],[2,3],[3,1]]; EP=[[1],[2],[3]]; d=1; # Triangle graph $r = 3$; three endpoints $m = 3$; dimension $d = 1$			
Instruction	Order	Cumulant output	Connected non-flat partitions
c(1,d,G,EP,mu,H)	1st	$\frac{\lambda^3}{4}$	1
c(2,d,G,EP,mu,H)	2nd	$\left(\frac{\sqrt{3}}{8}+\frac{3}{8}\right)\lambda^5+\left(\frac{2\sqrt{105}}{35}+\frac{\sqrt{3}}{5}+\frac{3}{4}\right)\lambda^4 + \left(\frac{3\sqrt{35}}{35}+\frac{\sqrt{2}}{5}+\frac{1}{4}\right)\lambda^3$	33

Table 7. *Computation times and counts of connected non-flat* vs. *all partitions in* $\Pi([n]\times[3])$.

Order n	3 blocks	4 blocks	5 blocks	6 blocks	7 blocks	Total	$\Pi([n]\times[3])$	Comp. time
1st	1	0	0	0	0	1	5	1 s
2nd	6	18	9	0	0	33	203	21 s
3rd	36	540	1,242	864	189	2,871	21,147	1 hour

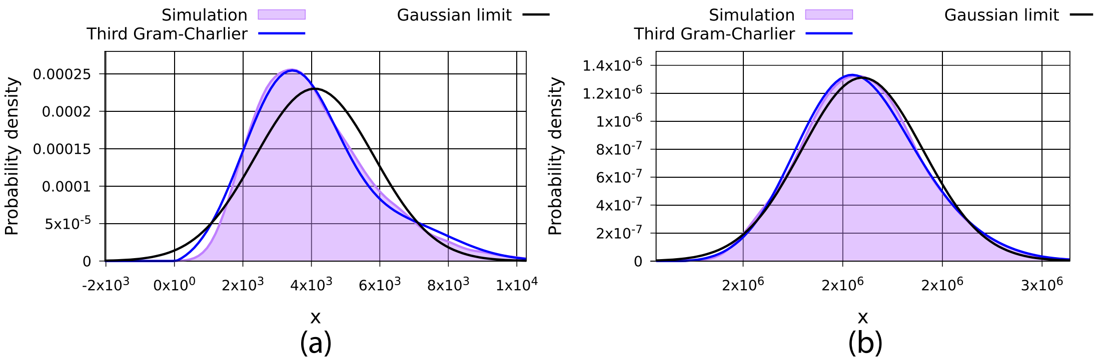

Figure 9. *Gram–Charlier density expansions* vs. *Monte Carlo density estimation. (a)* $\lambda = 50$. *(b)* $\lambda = 400$.

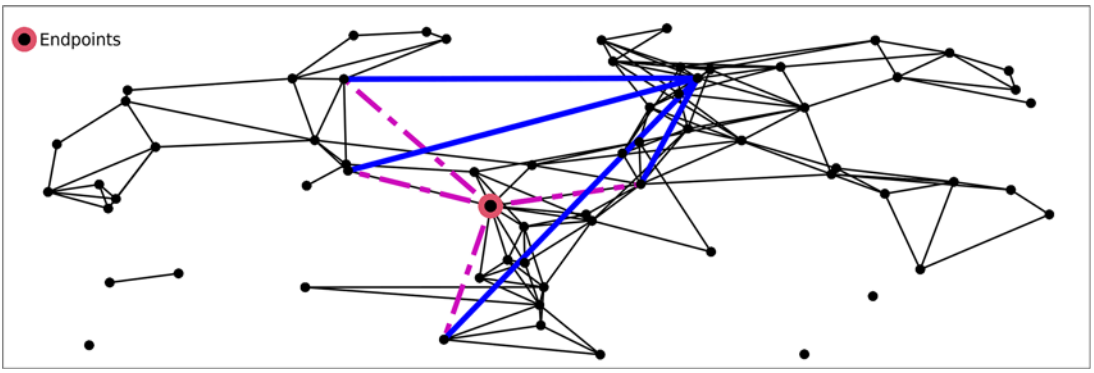

Figure 10. *Four trees with a single endpoint in dimension* $\mathrm{d}=2$.

Table 8. *First and second cumulants of the count of trees with one endpoint.*

G = [[1,2],[2,3],[2,4]]; EP=[[1,3,4]]; d=2;		# Tree $r=4$; single endpoint $m=1$; dimension $d=2$	
Instruction	Order	Cumulant output	Connected non-flat partitions
c(1,d,G,EP,mu,H)	1st	$\frac{\lambda^4}{12}$	1
c(2,d,G,EP,mu,H)	2nd	$\frac{41\lambda^7}{384}+\frac{99039\lambda^6}{165760}+\frac{232885\lambda^5}{175824}+\frac{37\lambda^4}{50}$	208

In Table 8, the location of the unique endpoint has no impact on cumulant expressions due to space homogeneity of the underlying Poisson point process.
The computation times presented in Table 9 are for dimension $d=2$.
In Figure 11, we plot the second cumulant of $N^G_{y_1,\ldots,y_m}$ and the third cumulant of $\widetilde{N}^G_{y_1,\ldots,y_m}$ *vs.* their Monte Carlo estimates in dimension $d=2$ with $y_1=y_2=y_3=0$.

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

Table 9. *Computation times and counts of connected non-flat* vs. *all partitions in* $\Pi([n] \times [4])$.

Order n	4 blocks	5 blocks	6 blocks	7 blocks	8 blocks	9 blocks	10 blocks	Total	$\Pi([n] \times [4])$	Comp. time
1st	1	0	0	0	0	0	0	1	15	1 s
2nd	24	96	72	15	0	0	0	208	4,140	2 m
3rd	576	13,824	50,688	59,904	29,952	6,912	640	162,496	4,213,597	40 hours

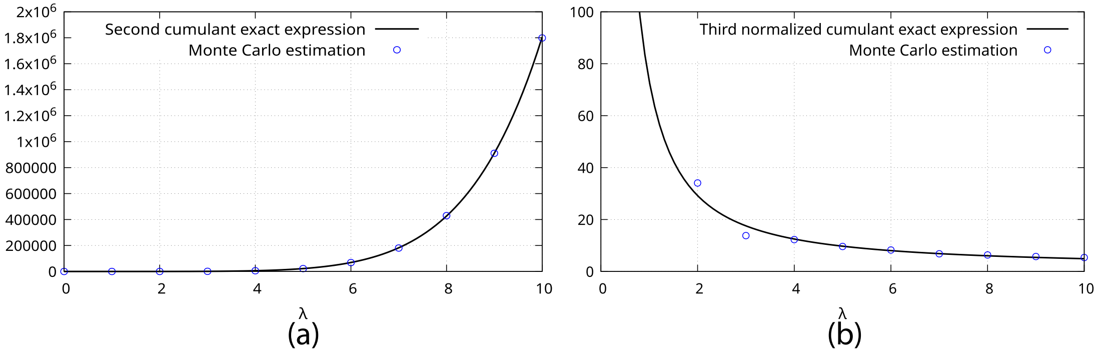

Figure 11. *Cumulant estimates. (a) Second cumulant. (b) Normalized third cumulant.*

Table 10. *Functions definitions.*

`load("cumulants_parallel.sage");` `load("jointcumulants.sage")`	# Loading the functions definitions
`λ,β= var("λ,β"); assume(β> 0)`	# Variable definitions
`def H(x,y,β): return exp(-β*(x-y)**2)`	# Connection function $H(x,y)$
`def mu(x,λ,β): return exp(-β*x**2)`	# Finite intensity measure $\mu(dx)$

Table 11. *Second (joint) moments of triangle counts* vs. *four-hop counts.*

G1 = [[1,2],[2,3],[3,1]]; G2 = [[1,2],[2,3],[3,4],[4,5]]; G2c = [[4,5],[5,6],[6,7],[7,8]]; G = [G1,G2c]; EP=[]; d=2;

Triangles G1 and four-hop G2; $r_1 = 3$, $r_2 = 5$; no endpoints $m=0$; dimension $d=2$

Instruction	Order	Cumulant output	Connected non-flat partitions
c(2,d,G1,EP,mu,H)	2nd	$\frac{3\lambda^5}{64}+\frac{6\lambda^4}{25}+\frac{3\lambda^3}{8}$	33
c(2,d,G2,EP,mu,H)	2nd	$\frac{7344738590701\lambda^9}{687218605505250}+\cdots$	1,545
jc(d,G,EP,mu,H)	2nd joint	$\frac{34409\lambda^7}{1537920}+\frac{9101145477\lambda^6}{55004486680}+\frac{10774977\lambda^5}{28148120}$	135

5.5. *Correlation of triangles* vs. *four-hop counts*

In this example, we run the joint cumulant code provided in Appendix E to compute the correlation of triangle and four-hop counts without endpoints, as a function of the intensity parameter λ. Here, μ is taken to be a finite measure as no endpoints are considered, that is we have EP=[ ] and $m=0$, and the SageMath code listed in Appendix E is run after loading the definitions of Table 10.
The closed-form expressions in Table 11 are expressed with $\beta := \pi$, in dimension $d=2$.
In Figure 12, we plot the second joint cumulant and correlation of (N^{G_1}, N^{G_2}) *vs.* their Monte Carlo estimates in dimension $d=1$.
The limit correlation as λ tends to infinity can be exactly estimated from Table 11 as

$$\frac{34409}{1537920}\sqrt{\frac{64}{3}\times\frac{687218605505250}{7344738590701}} \approx 0.999602.$$

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

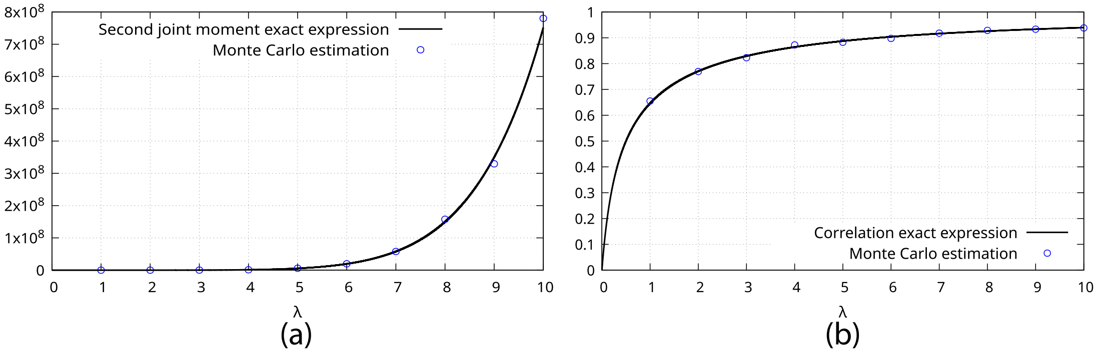

Figure 12. *Correlation and second joint cumulant estimates. (a) Second joint cumulant. (b) Correlation.*

Acknowledgments. This research is supported by the Ministry of Education, Singapore, under its Tier 2 Grant MOE-T2EP20120-0005. We thank Xueying Yang for essential contributions to the SageMath cumulant codes.

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

Supplementary material. To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0269964824000160.

References

[1] Bogdan, K., Rosiński, J., Serafin, G., & Wojciechowski, L. (2017). Lévy systems and moment formulas for mixed Poisson integrals. In *Stochastic analysis and related topics* (*Progress in probability*), vol. 72. Cham: Birkhäuser/Springer, pp. 139–164.
[2] Bollobás, B. (2001). *Random graphs* (*Cambridge studies in advanced mathematics*), vol. 73, 2nd ed. Cambridge: Cambridge University Press.
[3] Can, V.H. & Trinh, K.D. (2022). Random connection models in the thermodynamic regime: Central limit theorems for add-one cost stabilizing functionals. *Electronic Journal of Probability* 27: 1–40.
[4] Cramér, H. (1946). *Mathematical methods of statistics.* Princeton, NJ: Princeton University Press.
[5] Cunningham, W., Zuev, K., & Krioukov, D. (2017). Navigability of random geometric graphs in the universe and other spacetimes. *Scientific Reports* 7(1): 8699.
[6] Döring, H., Jansen, S., & Schubert, K. (2022). The method of cumulants for the normal approximation. *Probability Surveys* 19: 185–270.
[7] Fountoulakis, N. & Yukich, J.E. (2020). Limit theory for isolated and extreme points in hyperbolic random geometric graphs. *Electronic Journal of Probability* 25(2020): 51.
[8] Georgiou, O., Bocus, M.Z., Rahman, M.R., Dettmann, C.P., & Coon, J.P. (2015). Network connectivity in non-convex domains with reflections. *IEEE Communications Letters* 19(3): 427–430.
[9] Giles, A.P., Georgiou, O., & Carl, P.D. (2016). Connectivity of soft random geometric graphs over annuli. *Journal of Statistical Physics* 162(4): 1068–1083.
[10] Grote, J. & Thäle, C. (2018). Gaussian polytopes: A cumulant-based approach. *Journal of Complexity* 47: 1–41.
[11] Janson, S. (1988). Normal convergence by higher semiinvariants with applications to sums of dependent random variables and random graphs. *The Annals of Probability* 16(1): 305–312.
[12] Janson, S., Łuczak, T., & Rucinski, A. (2000). *Random graphs* (*Wiley-Interscience series in discrete mathematics and optimization*). New York: Wiley-Interscience, pp. xii+333.
[13] Kartun-Giles, A.P., Koufos, K., Lu, X., & Niyato, D. (2023). Two-hop connectivity to the roadside in a VANET under the random connection model. *IEEE Transactions on Vehicular Technology* 72(4): 5508–5512.
[14] Khorunzhiy, O. (2008). On connected diagrams and cumulants of Erdős–Rényi matrix models. *Communications in Mathematical Physics* 282(1): 209–238.
[15] Last, G., Nestmann, F., & Schulte, M. (2021). The random connection model and functions of edge-marked Poisson processes: Second order properties and normal approximation. *The Annals of Applied Probability* 31(1): 128–168.
[16] Liu, Q. & Privault, N. (2024). Normal approximation of subgraph counts in the random-connection model. *Bernoulli* 30(4): 3224–3250.
[17] Lukacs, E. (1955). Applications of Faàdi Bruno's formula in mathematical statistics. *The American Mathematical Monthly* 62: 340–348.

[18] Malyshev, V.A. & Minlos, R.A. (1991). *Gibbs random fields (Mathematics and its applications (Soviet Series)*, vol. 44. Dordrecht: Kluwer Academic Publishers Group.
[19] Mao, G., Zhang, Z., & Anderson, B.D.O. (2010). Probability of k-hop connection under random connection model. *IEEE Communications Letters* 14(11): 1023–1025.
[20] Ng, S.C., Zhang, W., Zhang, Y., Yang, Y., & Mao, G. (2011). Analysis of access and connectivity probabilities in vehicular relay networks. *IEEE Journal on Selected Areas in Communications* 29(1): 140–150.
[21] Peccati, G. & Taqqu, M. (2011). *Wiener chaos: Moments, cumulants and diagrams: A survey with computer implementation (Bocconi & Springer Series)*. Milan: Springer.
[22] Privault, N. (2012). Moments of Poisson stochastic integrals with random integrands. *Probability and Mathematical Statistics* 32(2): 227–239.
[23] Privault, N. (2019). Moments of k-hop counts in the random-connection model. *Journal of Applied Probability* 56(4): 1106–1121.
[24] Roch, S. (2023). Modern discrete probability – An essential toolkit. https://people.math.wisc.edu/~roch/mdp/roch-mdp-full.pdf. Last Accessed: 22 July 2023.
[25] Rudzkis, R., Saulis, L., & Statulevičius, V.A. (1978). A general lemma on probabilities of large deviations. *Lithuanian Mathematical Journal* 18(2): 99–116.
[26] Saulis, L. & Statulevičius, V.A. (1991). *Limit theorems for large deviations (Mathematics and its applications (Soviet Series))*, vol. 73. Dordrecht: Kluwer Academic Publishers Group.
[27] Snijders, T.A.B. & Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification* 14(1): 75–100.
[28] Ta, X., Mao, G., & Anderson, B.D.O. (2007). On the probability of k-hop connection in wireless sensor networks. *IEEE Communications Letters* 11(8): 662–664.
[29] Zhang, W., Chen, Y., Yang, Y., Wang, X., Zhang, Y., Hong, X., & Mao, G. (2012). Multi-hop connectivity probability in infrastructure-based vehicular networks. *IEEE Journal on Selected Areas in Communications* 30(4): 740–747.

Appendix A. Multivariate moment and cumulant formulae

In this section, we prove an extension of Proposition 3.4 for the joint moments and cumulants of subgraph counts. The next definition extends Definition 3.1.

Definition A.1. *Given $r_1, \dots, r_n \geq 1$, we set*

$$\pi_i = \{(i,1), \dots, (i,r_i)\}, \quad i = 1, \dots, n,$$

and $\pi := \{\pi_1, \dots, \pi_n\}$.

(i) A set partition $\sigma \in \Pi(\pi_1 \cup \cdots \cup \pi_n)$ is connected if $\sigma \vee \pi = \widehat{1}$.
(ii) A set partition $\sigma \in \Pi(\pi_1 \cup \cdots \cup \pi_n)$ is non-flat if $\sigma \wedge \pi = \widehat{0}$.

We let $\Pi_{\widehat{1}}(\pi_1 \cup \cdots \cup \pi_n)$ denote the collection of all connected partitions of $\pi_1 \cup \cdots \cup \pi_n$.

In what follows, every partition $\rho \in \Pi(\pi_1 \cup \cdots \cup \pi_n)$ will be arranged into a diagram denoted by $\Gamma(\rho, \pi)$, by arranging $\pi_1, \dots, \pi_n$ into n rows and connecting together the elements of every block of ρ, see Figure A1 for two illustrations with $n = 5$, $(r_1, r_2, r_3, r_4, r_5) = (3, 2, 4, 3, 4)$.
Definition A.2 extends [16, Def. 2.4] to the multivariate setting.

Definition A.2.

(i) Given $\rho \in \Pi(\pi_1 \cup \cdots \cup \pi_n)$, we let σ_ρ be the partition of $[n]$ defined by the condition

$$\rho \vee \pi = \left\{ \bigcup_{i \in b} \pi_i : b \in \sigma_\rho \right\}.$$

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

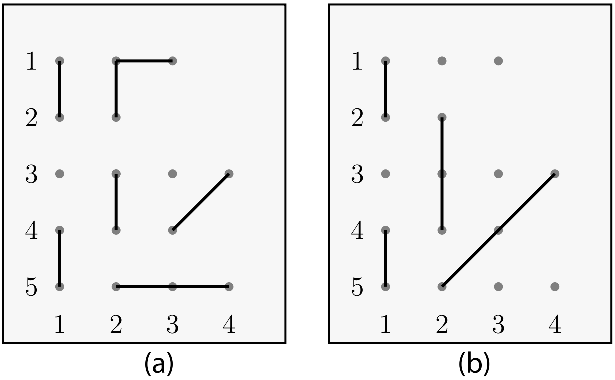

Figure A1. *Two examples of partition diagrams. (a) Non-connected partition diagram* $\Gamma(\rho, \pi)$*. (b) Connected partition diagram* $\Gamma(\rho, \pi)$*.*

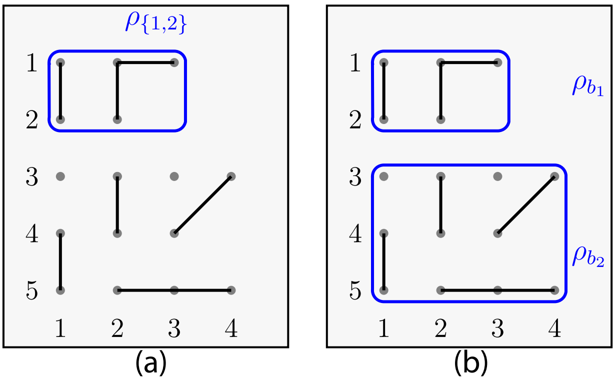

Figure A2. *Diagram* $\Gamma(\rho, \pi)$ *and splitting of the partition* ρ *with* $\rho \vee \pi = \{\pi_1 \cup \pi_2, \pi_3 \cup \pi_4 \cup \pi_5\}$*. (a) Connected subpartition* $\rho_{\{1,2\}}$*. (b) Splitting* ρ *into connected subpartitions* ρ_{b1}, ρ_{b2}*.*

(ii) For any non-empty set $b \subset [n]$*, we let*

$$\rho_b := \left\{ c \in \rho : c \subset \bigcup_{i \in b} \pi_i \right\}.$$

As an example, in Figure A2(a), when $b = \{1, 2\}$ we have

$$\rho_{\{1,2\}} = \big\{ \{(1,1), (2,1)\}, \{(1,2), (1,3), (2,2)\} \big\}.$$

We note that for $b \subset [n]$ we have $\pi_b = \{\pi_i : i \in b\}$, and any partition $\rho \in \Pi(\pi_1 \cup \cdots \cup \pi_n)$ can be split into subpartitions deduced from the connected components of $\Gamma(\rho, \pi)$, that is

$$\rho = \bigcup_{b \in \sigma_\rho} \rho_b,$$

as illustrated in Figure A2(b) with $b_1 = \{1, 2\}$, $b_2 = \{3, 4, 5\}$, and $\sigma_\rho = \{b_1, b_2\}$.

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

Definition A.3. *For $\sigma \in \Pi([n])$ we let $\Pi_\sigma(\pi_1 \cup \cdots \cup \pi_n)$ denote the collection of partitions $\rho \in \Pi(\pi_1 \cup \cdots \cup \pi_n)$ such that*

$$\rho \vee \pi = \left\{ \bigcup_{i \in b} \pi_i : b \in \sigma \right\}.$$

In particular, $\Pi_{\widehat{1}}(\pi_1 \cup \cdots \cup \pi_n)$ represents the set of connected partitions of $\pi_1 \cup \cdots \cup \pi_n$, and $\Pi_{\widehat{0}}(\pi_1 \cup \cdots \cup \pi_n)$ represents the partitions of $\pi_1 \cup \cdots \cup \pi_n$ that are finer than $\pi := \{\pi_1, \ldots, \pi_n\}$.

Given $F : \Pi'(\pi_1 \cup \cdots \cup \pi_n) \longrightarrow \mathbb{R}$, where $\Pi'(\pi_1 \cup \cdots \cup \pi_n)$ is the collection of all subpartitions of $\pi_1 \cup \cdots \cup \pi_n$, we define the mixed moments $\widehat{F} : 2^{[n]} \longrightarrow \mathbb{R}$ by

$$\widehat{F}(A) = \sum_{\rho \in \Pi(\cup_{i \in A} \pi_i)} F(\rho), \qquad A \subset [n], \tag{A.1}$$

cf. [18, p. 33]. The semi-invariants $C_F : 2^{[n]} \longrightarrow \mathbb{R}$ are defined by the induction formula $C_F(A) = \widehat{F}(A)$ when $|A| = 1$, and

$$C_F(A) = \widehat{F}(A) - \sum_{\substack{\{b_1,\ldots,b_k\} \in \Pi(A) \\ k \geq 2}} \prod_{i=1}^{k} C_F(b_i),$$

for $|A| > 1$, see Relation (16) page 33 of [18], where the sum is taken over all partitions $\sigma \in \Pi(A)$ such that $|\sigma| \geq 2$, that is

$$C_F(A) = \sum_{\rho \in \Pi(A)} (-1)^{|\rho|} (|\rho| - 1)! \prod_{b \in \rho} \widehat{F}(b), \tag{A.2}$$

see Relation (16') in [18]. The next proposition generalizes [16, Prop. 3.3] to the multivariate case.

Proposition A.4. *Suppose that F satisfies the connectedness factorization property*

$$F(\rho) = \prod_{b \in \sigma_\rho} F(\rho_b), \qquad \rho \in \Pi'(\pi_1 \cup \cdots \cup \pi_n). \tag{A.3}$$

Then, the semi-invariants are given by

$$C_F(A) = \sum_{\rho \in \Pi_{\widehat{1}}(\cup_{i \in A} \pi_i)} F(\rho), \qquad \emptyset \neq A \subset [n]. \tag{A.4}$$

Proof.

(i) It is clear that (A.4) holds when $|A| = 1$. When $|A| = 2$, taking $A = \{i,j\} \subset [n]$, $i \neq j$, we have

$$\begin{aligned} C_F(A) &= \widehat{F}(\{i,j\}) - C_F(\{i\}) C_F(\{j\}) \\ &= \sum_{\rho \in \Pi(\pi_i \cup \pi_j)} F(\rho) - \widehat{F}(\{i\}) \widehat{F}(\{j\}) \\ &= \sum_{\rho \in \Pi_{\widehat{1}}(\pi_i \cup \pi_j)} F(\rho) + \sum_{\rho \in \Pi_{\widehat{0}}(\pi_i \cup \pi_j)} F(\rho) - \left(\sum_{\rho_1 \in \Pi(\pi_i)} F(\rho_1) \right) \left(\sum_{\rho_2 \in \Pi(\pi_j)} F(\rho_2) \right). \end{aligned}$$

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

By splitting any $\rho \in \Pi_{\widehat{0}}(\pi_i \cup \pi_j)$ into two disjoint subpartitions according to Definition A.2, that is

$$\rho = \rho_{\{i\}} \cup \rho_{\{j\}},$$

together with the factorization property (A.3), we find

$$\begin{aligned}\sum_{\rho\in\Pi_{\widehat{0}}(\pi_i\cup\pi_j)} F(\rho) &= \sum_{\substack{\rho\in\Pi_{\widehat{0}}(\pi_i\cup\pi_j)\\ \rho=\rho_{\{i\}}\cup\rho_{\{j\}}}} F(\rho_{\{i\}})F(\rho_{\{j\}})\\ &= \left(\sum_{\rho_1\in\Pi(\pi_i)} F(\rho_1)\right)\left(\sum_{\rho_2\in\Pi(\pi_j)} F(\rho_2)\right),\end{aligned}$$

which shows (A.4).

(ii) Next, suppose that (A.4) holds for any $A \subset [n]$ with $|A| \le l \le n-1$. Let $A \subset [n]$ be a subset of $[n]$ with $|A| = l+1$. We have

$$\begin{aligned}\widehat{F}(A) &= \sum_{\rho\in\Pi(\cup_{i\in A}\pi_i)} F(\rho)\\ &= \sum_{\substack{\sigma=\{b_1,\ldots,b_k\}\in\Pi(A)\\ k\ge 1}} \sum_{\substack{\rho\in\Pi(\cup_{i\in A}\pi_i)\\ \rho\vee\pi_A=\{\cup_{i\in b_j}\pi_i\}_{j=1}^k}} F(\rho)\\ &= \sum_{\substack{\sigma=\{b_1,\ldots,b_k\}\in\Pi(A)\\ k\ge 1}} \sum_{\substack{\rho\in\Pi(\cup_{i\in A}\pi_i)\\ \rho\vee\pi_A=\{\cup_{i\in b_j}\pi_i\}_{j=1}^k}} \prod_{j=1}^k F(\rho_{b_j})\\ &= \sum_{\substack{\sigma=\{b_1,\ldots,b_k\}\in\Pi(A)\\ k\ge 1}} \prod_{j=1}^k \sum_{\substack{\rho_j\in\Pi(\cup_{i\in b_j}\pi_i)\\ \rho_j\vee\pi_{b_j}=\widehat{1}}} F(\rho_{b_i})\\ &= \sum_{\substack{\sigma=\{b_1,\ldots,b_k\}\in\Pi(A)\\ k\ge 1}} \prod_{j=1}^k \sum_{\rho_j\in\Pi_{\widehat{1}}(\cup_{i\in b_j}\pi_i)} F(\rho_{b_i})\\ &= \sum_{\rho\in\Pi_{\widehat{1}}(\cup_{i\in A}\pi_i)} F(\rho) + \sum_{\substack{\{b_1,\ldots,b_k\}\in\Pi(A)\\ k\ge 2}} \prod_{j=1}^k C_F(b_j),\end{aligned}$$

where the last equality follows from the induction hypothesis (A.4) when $|A| \le l$. The proof is completed by subtracting the last term from both sides.

□

Given $n \ge 1$ and $f^{(i)} : (\mathbb{R}^d)^{r_i} \to \mathbb{R}$, $i = 1, \ldots, n$, measurable functions, we let

$$\left(\bigotimes_{i=1}^n f^{(i)}\right)(x_{1,1}, \ldots, x_{1,r_1}, \ldots, x_{n,1}, \ldots, x_{n,r_n}) := \prod_{i=1}^n f^{(i)}(x_{i,1}, \ldots, x_{i,r_i}).$$

For $\rho \in \Pi(\pi_1 \cup \cdots \cup \pi_n)$, we also denote by $\left(\bigotimes_{i=1}^n f^{(i)}\right)_\rho : (\mathbb{R}^d)^{|\rho|} \to \mathbb{R}$ the function obtained by equating any two variables whose indexes belong to a same block of ρ. We refer to [1, Thm. 3.1] for the next result.

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

Proposition A.5. *Let $n \geq 1$, $r_1, \ldots, r_n \geq 1$, and let $f^{(i)} : (\mathbb{R}^d)^{r_i} \to \mathbb{R}$ be a sufficiently integrable measurable function for $i = 1, \ldots, n$. We have*

$$\mathbb{E}\left[\prod_{i=1}^{n} \sum_{(x_1,\ldots,x_{r_i})\in\eta^{r_i}} f^{(i)}(x_1,\ldots,x_{r_i})\right] = \sum_{\rho\in\Pi(\pi_1\cup\cdots\cup\pi_n)} \lambda^{|\rho|} \int_{(\mathbb{R}^d)^{|\rho|}} \left(\bigotimes_{i=1}^{n} f^{(i)}\right)_{\rho} (\mathbf{x})\, \mu^{\otimes|\rho|}(d\mathbf{x}),$$

Proposition A.5 can be specialized as follows.

Corollary A.6. *Let $r_i \geq 2$, $i = 1, \ldots, n$, and consider $f^{(i)} : (\mathbb{R}^d)^{r_i} \to \mathbb{R}$ measurable functions that vanish on diagonals, that is $f^{(i)}(x_1, \ldots, x_{r_i}) = 0$ whenever $x_k = x_l$ for some $1 \leq k \neq l \leq r_i$, $i = 1, \ldots, n$. We have*

$$\mathbb{E}\left[\prod_{i=1}^{n} \sum_{(x_1,\ldots,x_{r_i})\in\eta^{r_i}} f^{(i)}(x_1,\ldots,x_{r_i})\right] = \sum_{\substack{\rho\in\Pi(\pi_1\cup\cdots\cup\pi_n)\\ \rho\wedge\pi=\widehat{0}\\ (\text{non-flat})}} \lambda^{|\rho|} \int_{(\mathbb{R}^d)^{|\rho|}} \left(\bigotimes_{i=1}^{n} f^{(i)}\right)_{\rho} (\mathbf{x})\, \mu^{\otimes|\rho|}(d\mathbf{x}). \quad \text{(A.5)}$$

For $i = 1, \ldots, n$, let $M_i \subset \{1, \ldots, m\}$, $r_i \geq 2$, and let $G_i = (V_{G_i}, E_{G_i})$ be a connected graph with edge set E_{G_i} and vertex set of the form $V_{G_i} = \left(v_1^{(i)}, \ldots, v_{r_i}^{(i)}; \{w_j^{(i)}\}_{j\in M_i}\right)$, such that

(i) the subgraph G_i induced by G_i on $\{v_1^{(i)}, \ldots, v_{r_i}^{(i)}\}$ is connected, and
(ii) the endpoint vertices $\{w_j^{(i)}\}_{j\in M_i}$ are not adjacent to each other in G_i,

and let $G := \{G_1, \ldots, G_n\}$. In Definition A.7, for every $\rho \in \Pi(\pi_1 \cup \cdots \cup \pi_n)$ we build a graph structure induced by $(G_1, \ldots, G_n)$ on the diagram $\Gamma(\rho, \pi)$, analogous to [16, Def. 2.2].

Definition A.7. *Given $\rho \in \Pi(\pi_1 \cup \cdots \cup \pi_n)$ a partition of $\pi_1 \cup \cdots \cup \pi_n$, we let $\widetilde{\rho}_G$ denote the multigraph constructed as follows on $[m] \cup \pi_1 \cup \cdots \cup \pi_n$:*

(i) for all $j_1, j_2 \in [r_i]$, $j_1 \neq j_2$, and $i \in [n]$, an edge links (i, j_1) to (i, j_2) iff $\{v_{j_1}^{(i)}, v_{j_2}^{(i)}\} \in E_{G_i}$.
(ii) for all $(j, k) \in [r_i] \times M_i$ and $i \in [n]$, an edge links (k) to (i, j) iff $\{v_j^{(i)}, w_k^{(i)}\} \in E_{G_i}$;
(iii) for all $i_1, i_2 \in [n]$ and $(j_1, j_2) \in [r_{i_1}] \times [r_{i_2}]$, we merge any two nodes (i_1, j_1) and (i_2, j_2) if they belong to a same block in ρ.

In addition, we let ρ_G be the graph constructed from the multigraph $\widetilde{\rho}_G$ by removing any redundant edge in $\widetilde{\rho}_G$.

As in Section 3, the graph ρ_G forms a connected graph with $|\rho| + m$ vertices. Figure A3 presents two examples of multigraphs $\widetilde{\rho}_G$ and graphs ρ_G when G_1, G_2, G_3 are line graphs, G_4 is a triangle, and G_5 is a rectangle on a partition diagram $\Gamma(\rho, \pi)$ with no endpoints, that is $M_1 = \cdots = M_n = \emptyset$ here.
For each $i = 1, \ldots, n$ denote by $N_{M_i}^{G_i}$ the count of subgraphs in the random-connection model $G_H(\eta \cup \{y_j\}_{j\in M_i})$ with endpoint set $\{y_j\}_{j\in M_i}$, that is

$$N_{M_i}^{G_i} = \sum_{(x_1,\ldots,x_{r_i})\in\eta^{r_i}} f_{M_i}^{(i)}(x_1, \ldots, x_{r_i}),$$

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

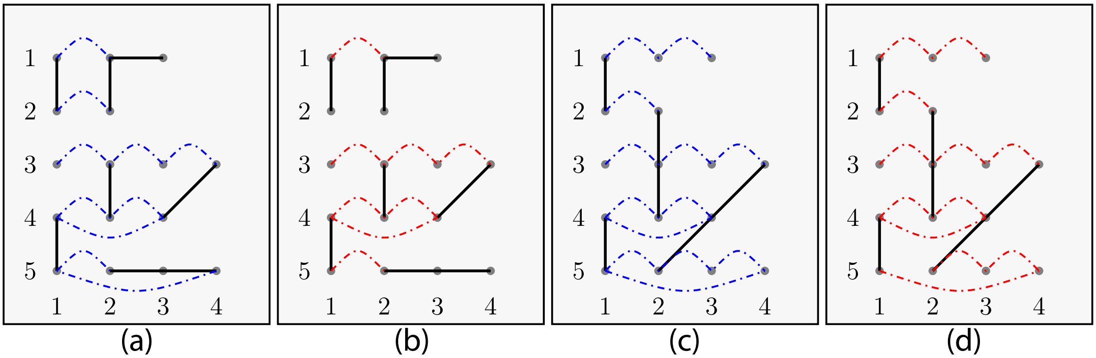

Figure A3. *Diagram $\Gamma(\rho, \pi)$, multigraph $\widetilde{\rho}_G$, and graph ρ_G. (a) Multigraph $\widetilde{\rho}_G$ in blue. (b) Graph ρ_G in red. (c) Multigraph $\widetilde{\rho}_G$ in blue. (d) Graph and ρ_G in red.*

where $f^{(i)}_{M_i} : (\mathbb{R}^d)^{r_i} \longrightarrow \{0, 1\}$ is the random function defined as

$$f^{(i)}_{M_i}(x_1, \ldots, x_{r_i}) := \prod_{\substack{1\le l\le r_i,\, j\in M_i \\ \{w^{(i)}_j, v^{(i)}_l\}\in E_{G_i}}} \mathbf{1}_{\{y_j \leftrightarrow x_l\}} \prod_{\substack{1\le k,l\le r_i \\ \{v^{(i)}_k, v^{(i)}_l\}\in E_{G_i}}} \mathbf{1}_{\{x_k \leftrightarrow x_l\}}, \qquad x_1, \ldots, x_{r_i} \in \mathbb{R}^d.$$

For $\rho = \{b_1, \ldots, b_{|\rho|}\} \in \Pi(\pi_1 \cup \cdots \cup \pi_n)$, we also let

$$\mathcal{A}^\rho_j := \{i \in [|\rho|] \ : \ \exists (s, k) \in b_i \text{s.t.} (v^{(s)}_k, w^{(s)}_j) \in E_{G_s}\}$$

denote the neighborhood of the vertex $(|\rho|+j)$ in $\rho_G, j = 1, \ldots, m$. The next proposition is a consequence of Relation (A.4) and Corollary A.6.

Proposition A.8. *Let $N^{G_i}_{M_i}$ be subgraph counts in the random-connection model $G_H(\eta \cup \{y_1, \ldots, y_m\})$ as defined above, for $i = 1, \ldots, n$. We have*

$$\mathbb{E}\left[\prod_{i=1}^n N^{G_i}_{M_i}\right] = \sum_{\substack{\rho\in\Pi(\pi_1\cup\cdots\cup\pi_n) \\ \rho\wedge\pi=\widehat{0} \\ \text{(non-flat)}}} \lambda^{|\rho|} \int_{(\mathbb{R}^d)^{|\rho|}} \prod_{\substack{1\le j\le m \\ i\in\mathcal{A}^\rho_j}} H(x_i, y_j) \prod_{(k,l)\in E_{\rho_G}} H(x_k, x_l)\, \mu(\mathrm{d}\mathbf{x}), \tag{A.6}$$

and joint cumulant

$$\kappa(N_{G_1}, \ldots, N_{G_n}) = \sum_{\substack{\rho\in\Pi_{\widehat{1}}(\pi_1\cup\cdots\cup\pi_n) \\ \rho\wedge\pi=\widehat{0} \\ \text{(non-flat connected)}}} \lambda^{|\rho|} \int_{(\mathbb{R}^d)^{|\rho|}} \prod_{\substack{1\le j\le m \\ i\in\mathcal{A}^\rho_j}} H(x_i, y_j) \prod_{(k,l)\in E_{\rho_G}} H(x_k, x_l)\, \mu(\mathrm{d}\mathbf{x}). \tag{A.7}$$

Proof. The moment identity (A.6) is obtained by taking expectation on both sides of (A.5) in Corollary A.6 and using the relation $H(x, y) = \mathbb{E}[\mathbf{1}_{\{x\leftrightarrow y\}}]$, $x, y \in \mathbb{R}^d$. Next, we note that the connectedness factorization property (A.3) is satisfied by

$$F(\rho) := \lambda^{|\rho|} \int_{(\mathbb{R}^d)^{|\rho|}} \prod_{\substack{1\le j\le m \\ i\in\mathcal{A}^\rho_j}} H(x_i, y_j) \prod_{(k,l)\in E_{\rho_G}} H(x_k, x_l)\, \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_{|\rho|}),$$

$\rho \in \Pi(\pi_1 \cup \cdots \cup \pi_n)$, hence (A.7) follows from Relations (A.1), (A.4), (A.6), and the classical cumulant-moment relationship (A.2), see, for example, Relation (3.3) in [17]. □

The cumulant formula of Proposition A.8 is implemented in the code listed in Appendix E.

Appendix B. Cumulant and factorial moment estimates

The following result can be found in [26, Cor. 2.1] or [6, Thm. 2.4].

Lemma B.1. *Let $\{X_\lambda\}$ be a family of random variables with moments of all orders, mean zero and unit variance for all $\lambda > 0$. Suppose that for all $j \geq 3$ and sufficiently large λ, the cumulant of order j of X_λ is bounded by*

$$|\kappa_j(X_\lambda)| \leq \frac{(j!)^{1+\gamma}}{(\Delta_\lambda)^{j-2}}$$

where $\gamma \geq 0$ is a constant independent of λ. Then we have the Berry–Esseen bound

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(X_\lambda \leq x) - \Phi(x)| \leq C_\gamma (\Delta_\lambda)^{-1/(1+2\gamma)},$$

for $C_\gamma > 0$ a constant depending only on γ.

We let $m_n(X) := \mathbb{E}[X(X-1)\cdots(X-n+1)]$ denote the factorial moments of order $n \geq 1$ of a discrete random variable X.

Proposition B.2. Corollary 1.13 in [2] *Assume that*

$$\lim_{n\to\infty} m_n(X) \frac{n^k}{n!} = 0, \qquad k \geq 0.$$

Then for any $n \geq 0$, we have

$$\mathbb{P}(X = n) = \frac{1}{n!} \sum_{i \geq 0} \frac{(-1)^i}{i!} m_{n+i}(X). \tag{B.1}$$

Appendix C. Gram–Charlier expansions

Let $\varphi(x) := e^{-x^2/2}/\sqrt{2\pi}$, $x \in \mathbb{R}$, denote the standard normal probability density function. In addition to the second order expansion Gaussian approximation

$$\phi_X^{(1)}(x) = \frac{1}{\sqrt{\kappa_2}} \varphi\left(\frac{x-\kappa_1}{\sqrt{\kappa_2}}\right) \tag{C.1}$$

for the probability density $\phi_X(x)$ function of a random variable X, higher order Gram–Charlier expansions of third and fourth order are given by

$$\phi_X^{(3)}(x) = \frac{1}{\sqrt{\kappa_2}} \varphi\left(\frac{x-\kappa_1}{\sqrt{\kappa_2}}\right)\left(1 + c_3 H_3\left(\frac{x-\kappa_1}{\sqrt{\kappa_2}}\right)\right) \tag{C.2}$$

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

and

$$\phi_X^{(4)}(x) = \frac{1}{\sqrt{\kappa_2}}\varphi\left(\frac{x-\kappa_1}{\sqrt{\kappa_2}}\right)\left(1 + c_3 H_3\left(\frac{x-\kappa_1}{\sqrt{\kappa_2}}\right) + c_4 H_4\left(\frac{x-\kappa_1}{\sqrt{\kappa_2}}\right) + c_6 H_6\left(\frac{x-\kappa_1}{\sqrt{\kappa_2}}\right)\right).$$

see Section 17.6 of [4], where

- $H_0(x) = 1$, $H_1(x) = x$, $H_3(x) = x^3 - 3x$, $H_4(x) = x^4 - 6x^2 + 3$, $H_6(x) = x^6 - 15x^4 + 45x^2 - 15$ are Hermite polynomials,
- the sequence c_3, c_4, c_5, c_6 is given from the cumulants $(\kappa_n)_{n\geq 1}$ of X as

$$c_3 = \frac{\kappa_3}{3!(\kappa_2)^{3/2}}, \quad c_4 = \frac{\kappa_4}{4!(\kappa_2)^2}, \quad c_5 = \frac{\kappa_5}{5!\kappa_5^{5/2}}, \quad c_6 = \frac{\kappa_6}{6!(\kappa_2)^3} + \frac{(\kappa_3)^2}{2(3!)^2(\kappa_2)^3},$$

where c_3 and c_4 are expressed from the skewness $\kappa_3/(\kappa_2)^{3/2}$ and the excess kurtosis $\kappa_4/(\kappa_2)^2$.

Appendix D. Cumulant code

The following code generates closed-form cumulant expressions via symbolic calculations in SageMath for any dimension $d\geq 1$, any connected subgraph G induced by G, and any set of endpoint connections represented as the sequence $\mathsf{EP} = [\mathsf{EP}_1, \ldots, \mathsf{EP}_m]$. When G has no endpoint ($m=0$) we have EP = [ ], however, in this case the measure μ should be finite, that is, the density function $\mathtt{mu}(\mathtt{x}, \lambda, \beta)$ should be integrable with respect to the Lebesgue measure on $\mathbb{R}^d$. The choice of SageMath for this implementation is due to its fast handling of symbolic integration via Maxima, which is significantly faster than the Python package Sympy. This code is also sped up by parallel processing that can distribute the load among different CPU cores. This SageMath code and the next one are available for download at https://github.com/nprivaul/random-connection.

```
from time import time
import datetime
import multiprocessing as mp

global cumulants

def partitions(points):
    if len(points) == 1:
        yield [ points ]
        return
    first = points[0]
    for smaller in partitions(points[1:]):
        for m, subset in enumerate(smaller):
            yield smaller[:m] + [[ first ] + subset]  + smaller[m+1:]
        yield [ [ first ] ] + smaller

def nonflat(partition,r):
    p = []
    for j in partition:
        seq = list(map(lambda x: (x-1)//r,j))
        p.append(len(seq) == len(set(seq)))
    return all(p)
```

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

```
def connected(partition,n,r):
    q = []; c = 0
    if n  == 1: return all([len(j)==1 for j in partition])
    for j in partition:
        jk = list(set(map(lambda x: (x-1)//r,j)))
        if(len(jk)>1):
            if c == 0:
                q = jk; c += 1
            elif(set(q) & set(jk)):
                d=[y for y in (q+jk) if y not in q]
                q = q + d
    return n == len(set(q))

def connectednonflat(n,r):
    points = list(range(1,n*r+1))
    randd = []
    for m, p in enumerate(partitions(points), 1):
        randd.append(sorted(p))
    cnfp = [e for e in randd if (connected(e,n,r) and nonflat(e,r))]
    for rou in range(r,(r-1)*n+2):
        rs = [d for d in cnfp if len(d)==rou]
        print("Connected␣non-flat␣partitions␣with",rou,"blocks:",len(rs))
    print("Connected␣non-flat␣set␣partitions:",len(cnfp))
    return cnfp

def graphs(G,EP,setpartition,n):
    r=len(set(flatten(G)));rhoG = []
    for j in range(n):
        for hop in G: rhoG.append([r*j+hop[0],r*j+hop[1]])
        for l in range(len(EP)):
            F=EP[l]
            for i in F: rhoG.append([j*r+i,n*r+l+1]);
    for i in setpartition:
        if(len(i)>1):
            b = []
            for j in rhoG:
                b.append([i[0] if ele in i else ele for ele in j])
            rhoG = b
    for i in rhoG: i.sort()
    return rhoG

def inner(n,d,G,EP,mu,H,setpartition,z,r):
    rhoG=graphs(G,EP,setpartition,n)
    for ll in range(len(EP)+1):
        for l in range(1,d+1): z[d*(n*r+ll)+l] = var(str(y)+str(ll)+str('_')+str(l
        ))
    for key in range(1,n*r+1):
        for l in range(1,d+1): z[key*d+l] = var(str(x)+str(key)+str(x)+str(l))
    edgesrhoG = [i for n, i in enumerate(rhoG) if i not in rhoG[:n]]
    vertrhoG = set(flatten(edgesrhoG));
    for ll in range(len(EP)): vertrhoG.remove(n*r+ll+1);
    strr = 'λ*'*len(vertrhoG)
    for i in vertrhoG:
        for l in range(1,d+1): strr = '*mu({},{},{})'.format(z[i*d+l], λ, β) + strr
        for l in range(1,d+1): strr = strr + ').integrate({},-infinity,+infinity)'
        .format(z[i*d+l])
    for i in edgesrhoG:
        for l in range(1,d+1): strr = '*H({},{},{})'.format(z[i[0]*d+l],z[i[1]*d+l
        ], β) + strr
    strr = '('*len(vertrhoG)*d+strr[1:]
    return eval(preparse(strr))

def collect_result(result):
    global cumulants
    global iii
    global tim
    iii=iii+1;
    if (mod(iii,100)==0):
        tim=(time()-t_start2)*(lencnfp-iii)/iii/60
        print('[%d]\r'%(iii),'Est.␣remaining␣time␣(minutes):%d'%(tim),end="")
    cumulants+=result
```

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

```
def c(n,d,G,EP,mu,H):
    global cumulants
    global iii
    global t_start2
    t_start2 = time()
    d_start2 = datetime.datetime.now()
    r=len(set(flatten(G)));
    x,y=var("x,y")
    cumulants = 0; iii = 0
    z = dict(enumerate([str(x)+str(key)+str(x)+str(l) for key in range(0,n*r+1)
        for l in range(1,d+1)], start=1))
    global lencnfp
    cnfp=connectednonflat(n,r)
    lencnfp=len(cnfp)
    pool = mp.Pool(4) # pool = mp.Pool(mp.cpu_count())
    for setpartition in cnfp:
        pool.apply_async(func = inner, args=(n,d,G,EP,mu,H,setpartition,z,r),
            callback=collect_result)
    pool.close()
    pool.join()
    print("\n");
    d_end2 = datetime.datetime.now()
    print("Runtime␣is",(d_end2-d_start2))
    return cumulants._sympy_()
```

Appendix E. Joint cumulant code

The following code generates closed-form joint cumulant expressions via symbolic calculations in SageMath for any dimension $d \geq 1$, any sequence $(\mathsf{G}_1, \ldots, \mathsf{G}_n)$ of connected subgraphs induced by $(G_1, \ldots, G_n)$. As above, the endpoint connections are represented using represented as the sequence $\mathsf{EP} = [\mathsf{EP}_1, \ldots, \mathsf{EP}_m]$ where EP_i denotes the set of vertices of $\mathsf{G}_1 \cup \cdots \cup \mathsf{G}_n$ which are attached to the ith endpoint, $i = 1, \ldots, m$.

```
def jpartitions(points):
    if len(points) == 1:
        yield [ points ]
        return
    first = points[0]
    for smaller in jpartitions(points[1:]):
        for m, subset in enumerate(smaller):
            yield smaller[:m] + [[ first ] + subset]  + smaller[m+1:]
        yield [ [ first ] ] + smaller

def jnonflat(partition,rr):
    n=len(rr); p = []
    for j in partition:
        for i in range(n):
            j2 = [l for l in j if l > sum(rr[0:i]) and l<=sum(rr[0:(i+1)])]
            p.append(len(j2) <= 1)
    return all(p)

def jconnected(partition,rr):
    n=len(rr); q = []; c = 0;
    if n  == 1: return True
    for j in partition:
        jk = [i for i in range(n) if len([l for l in j if l > sum(rr[0:i]) and l<=
            sum(rr[0:(i+1)])])>=1]
        if(len(jk)>1):
            if c == 0:
                q = jk; c += 1
            elif(set(q) & set(jk)):
                d=[y for y in (q+jk) if y not in q]
                q = q + d
    return n == len(q)
```

https://doi.org/10.1017/S0269964824000160 Published online by Cambridge University Press

```
def jconnectednonflat(rr):
    n=len(rr);
    points = list(range(1,sum(rr)+1))
    randd = []
    for m, p in enumerate(jpartitions(points), 1): randd.append(sorted(p))
    for rou in range(min(rr),sum(rr)-n+2):
        rs = [d for d in randd if (jnonflat(d,rr) and len(d)==rou)]
        rss = [e for e in rs if jconnected(e,rr)]
        print("Connected␣non-flat␣partitions␣with",rou,"blocks:",len(rss))
    cnfp = [e for e in randd if (jconnected(e,rr) and jnonflat(e,rr))]
    print("Connected␣non-flat␣set␣partitions:",len(cnfp))
    return cnfp

def jgraphs(G,EP,setpartition):
    rr=[len(set(flatten(g))) for g in G];
    n=len(G); rhoG = []
    ee=[len(set(flatten(e))) for e in EP];
    for j in range(n):
        for hop in G[j]: rhoG.append([hop[0],hop[1]])
        for l in range(len(EP)):
            F=EP[l]
            for i in F: rhoG.append([i,sum(rr)+l+1]);
    for i in setpartition:
        if(len(i)>1):
            b = []
            for j in rhoG:
                b.append([i[0] if ele in i else ele for ele in j])
            rhoG = b
    for i in rhoG: i.sort()
    return rhoG

def jc(d,G,EP,mu,H):
    rr=[len(set(flatten(g))) for g in G];
    if(sum(rr)!=len(set(flatten(G)))):
        print("Wrong␣G␣format");
        return 0
    n=len(G);
    ee=[len(set(flatten(e))) for e in EP];
    x,y=var("x,y")
    jcumulants = 0; ii=0
    z = dict(enumerate([str(x)+str(key)+str(x)+str(l) for key in range(1,sum(rr)
       +1) for l in range(1,d+1)], start=1))
    cnfp=jconnectednonflat(rr)
    for setpartition in cnfp:
        ii=ii+1;print('[%d/%d]\r'%(ii,len(cnfp)),end="")
        rhoG=jgraphs(G,EP,setpartition)
        for j in range(n):
            m=len(EP);
            for l in range(m+1):
                for ld in range(1,d+1): z[d*(sum(rr)+l)+ld] = var(str(y)+str(l)+
                    str('_')+str(ld))
        for key in range(1,sum(rr)+1):
            for l in range(1,d+1): z[key*d+l] = var(str(x)+str(key)+str(x)+str(l))
        edgesrhoG = [i for n, i in enumerate(rhoG) if i not in rhoG[:n]]
        vertrhoG = set(flatten(edgesrhoG));
        m=len(EP);
        for l in range(m): vertrhoG.remove(sum(rr)+l+1);
        strr = 'λ*'*len(vertrhoG)
        for i in vertrhoG:
            for l in range(1,d+1): strr = '*mu({},{},{})'.format(z[i*d+l], λ, β) +
               strr
            for l in range(1,d+1): strr = strr + ').integrate({},-infinity,+
               infinity)'.format(z[i*d+l])
        for i in edgesrhoG:
            for l in range(1,d+1): strr = '*H({},{},{})'.format(z[i[0]*d+l],z[i
               [1]*d+l], β) + strr
        strr = '('*len(vertrhoG)*d+strr[1:]
        jcumulants += eval(preparse(strr))
    print("\n");
    jcumulants = simplify(jcumulants).canonicalize_radical().maxima_methods().
       rootscontract().simplify()
    return jcumulants._sympy_()
```

Cite this article: Liu Q. and Privault N. (2025). Graph connectivity with fixed endpoints in the random-connection model. *Probability in the Engineering and Informational Sciences* 39(3): 370–396. https://doi.org/10.1017/S0269964824000160