Bernoulli **30**(4), 2024, 3224–3250
https://doi.org/10.3150/23-BEJ1712

Normal approximation of subgraph counts in the random-connection model

QINGWEI LIU[a] and NICOLAS PRIVAULT[b]

Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371, Singapore, [a]*qingwei.liu@ntu.edu.sg,* [b]*nprivault@ntu.edu.sg*

This paper derives normal approximation results for subgraph counts written as multiparameter stochastic integrals in a random-connection model based on a Poisson point process. By combinatorial arguments we express the cumulants of general subgraph counts using sums over connected partition diagrams, after cancellation of terms obtained by Möbius inversion. Using the Statulevičius condition, we deduce convergence rates in the Kolmogorov distance by studying the growth of subgraph count cumulants as the intensity of the underlying Poisson point process tends to infinity. Our analysis covers general subgraphs in the dilute and full random graph regimes, and tree-like subgraphs in the sparse random graph regime.

Keywords: Cumulant method; Kolmogorov distance; normal approximation; Poisson point process; random-connection model; random graphs; subgraph count

1. Introduction

This paper treats the asymptotic behavior of random subgraph counts in the random-connection model, which is used to model physical systems in e.g. wireless networks, complex networks, and statistical mechanics. Our approach relies on the study of cumulant growth rates as the intensity of the underlying Poisson point process tends to infinity.

The distributional approximation of subgraph counts has attracted significant interest in the random graph literature. In Ruciński (1988), conditions for the asymptotic normality of renormalized subgraph counts have been obtained in the Erdős-Rényi random graph model Erdős and Rényi (1959), Gilbert (1959). In (Penrose, 2003, Chapter 3), non-quantitative central limit theorems have been obtained for subgraph and component counts on random geometric graphs. Rates of normal convergence with respect to the Kolmogorov distance have been obtained for certain random functionals on random geometric graphs in Schulte (2016) using Poisson U-statistics, see also Lachièze-Rey, Schulte and Yukich (2019) for the use of stabilizing functionals.

In the Erdős-Rényi setting, those results have been made more precise in Barbour, Karoński and Ruciński (1989) by the derivation of convergence rates in the Wasserstein distance via the Stein method. They have also been strengthened in the Kolmogorov distance for triangle counts in Krokowski, Reichenbachs and Thäle (2017), and for general subgraph counts in Privault and Serafin (2020). The case of triangles has also been treated in Röllin (2022) by the Stein-Tikhomirov method, which has been extended to general subgraphs in Eichelsbacher and Rednoß (2023). In Khorunzhiy (2008), the counts of line (X-model) and cycles (Y-model) in discrete Erdős-Rényi models have been analyzed via the asymptotic behavior of their cumulants.

The random-connection model is a natural generalization of the Erdős-Rényi random graph in which vertices are randomly located and can be connected with location-dependent probabilities $H(x,y) \in [0,1]$. Obtaining normal approximation error bounds in the random-connection model with a general $[0,1]$-valued random connection function is more difficult due to the additional layer of complexity coming from the randomness of vertex locations.

Regarding convergence rates, in Last, Nestmann and Schulte (2021), a central limit theorem and Berry-Esseen convergence rates have been presented and applied to the number of components isomorphic to a given finite connected graph in the random-connection model, together with a study of first moments and covariances. In Zhang (2022), Berry-Esseen convergence rates have been obtained for subgraph counts in the binomial random-connection (graphon) model. However, those results do not cover the case of general subgraph counting in the Poisson random-connection model.

The Malliavin-Stein machinery on Poisson space Last, Peccati and Schulte (2016) has been applied to in Last, Nestmann and Schulte (2021) the numbers of components isomorphic to a given graph in the random-connection model, using edge marked Poisson processes. Recently, a central limit theorem has been derived in Can and Trinh (2022) for the counts of induced subgraphs in the random-connection model using the edge-marking structure of Last, Nestmann and Schulte (2021), under a weak stabilizing condition originating from Penrose and Yukich (2001). However, as pointed out in Remark 2.5-(i) of Can and Trinh (2022), no convergence rates are derived by this method, as the strong stabilization condition of Lachièze-Rey, Schulte and Yukich (2019), Penrose and Yukich (2005) is not satisfied by general functionals when the connection function $H(x,y)$ is $(0,1)$-valued. On the other hand, the cumulant method, combined with the use of partition diagrams, enables us to establish a quantitative central limit theorem for functionals in the random-connection model.

In this paper, we derive normal approximation rates under a mild condition on the connection function $H(x,y)$ of the random-connection model, by deriving growth rates of cumulants written as sums over connected partitions, see Propositions 6.3 and 6.5. Related cumulant bounds have been obtained in the Erdős-Rényi model, cf. Proposition 10.1.2 in Féray, Méliot and Nikeghbali (2016). However, to the best of our knowledge, this is the first time that the normal approximation of subgraph counts with convergence rates is established in the random-connection model.

In comparison with Khorunzhiy (2008), which also uses the cumulant method, we obtain convergence rates in the Kolmogorov distance and our results are not restricted to line and cycle graphs, as they cover more general subgraphs, see Corollaries 7.1-7.2. Furthermore, various random graph regimes are discussed. In addition, we show in Section 8 that our approach can be specialized to derive Kolmogorov rates for subgraph counting in the setting of random geometric graphs, see Corollary 8.4.

A number of probabilistic conclusions can be derived from the behavior of cumulants of random variables using the Statulevičius condition, including convergence rates in the Kolmogorov distance and moderate deviation principles, see Saulis and Statulevičius (1991), Döring and Eichelsbacher (2013), Döring, Jansen and Schubert (2022). In stochastic geometry, the cumulant method has also been applied to Poisson cylinder processes Heinrich and Spiess (2009), and to the volumes of simplices in Poisson-Delaunay tessellations Gusakova and Thäle (2021), to the Boolean model Heinrich (2007), and to random m-dependent fields Götze, Heinrich and Hipp (1995). In Grote and Thäle (2018a,b), this method has been used to derive concentration inequalities, normal approximation with error bounds, and moderate deviation principles for random polytopes.

Given μ a finite diffuse measure on $\mathbb{R}^d$, we consider a random-connection model based on an underlying Poisson point process Ξ on $\mathbb{R}^d$ with intensity of the form $\lambda\mu(\mathrm{d}x)$, in which any two vertices x, y in Ξ are connected with the probability $H_\lambda(x,y) := c_\lambda H(x,y) \in [0,1]$, where H_λ is the connection function of the model. Here, we investigate the limiting behavior of the count N_G of a given subgraph G as the intensity λ of the underlying Poisson point process on $\mathbb{R}^d$ tends to infinity. To this end, we use the combinatorics of the cumulants $\kappa_n(N_G)$ based on moment expressions obtained in Privault (2019) for multiparameter stochastic integrals in the random-connection model.

Using partition diagrams and dependency graph arguments, we start by showing in Proposition 3.3 that the (virtual) cumulants of a random functional admitting a certain connectedness factorization property (8) can be expressed as sums over connected partition diagrams, generalizing Lemma 2 in Malyshev and Minlos (1991). A related result has been obtained in Jansen (2019) in the particular case of two-parameter Poisson stochastic integrals, in relation to cluster expansions for Gibbs point processes in statistical mechanics. In Proposition 4.2, we apply Proposition 3.3 to express the cumulants of multiparameter stochastic integrals, for which this factorization property can be checked from the moment formulas for multiparameter stochastic integrals computed in Proposition 1.1.

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 Khorunzhiy (2008). We work under a mild condition (17) which is satisfied by e.g. any translation-invariant continuous connection function $H : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ non vanishing at 0, such as the Rayleigh connection function given by $H(x,y) = e^{-\beta\|x-y\|^2}$, $x,y \in \mathbb{R}^d$, for some $\beta > 0$.

For our analysis of cumulant behavior we identify the leading terms in the sum (16) over connected partition diagrams. When G is a connected graph with $|V(G)| = r$ vertices, satisfying Assumption 6.1 in the dilute regime (18) with $\lambda^{-1/\zeta} \ll c_\lambda \leq K$, where $\zeta \geq 1$ is defined in (19), the dominant terms are given by connected partition diagrams with the highest number of blocks, see also Privault (2022) in the case of k-hop counting on the line. In Proposition 6.3 this yields the cumulant bounds

$$(n-1)!c_\lambda^{n|E(G)|}(K_1\lambda)^{1+(r-1)n} \leq \kappa_n(N_G) \leq n!^r c_\lambda^{n|E(G)|}(K_2\lambda)^{1+(r-1)n}, \quad \lambda \geq 1,$$

for some constants K_1, $K_2 > 0$ independent of $\lambda, n \geq 1$, where $E(G)$ denotes the set of edges of G. From the Statulevičius condition (41) below, see Döring, Jansen and Schubert (2022), Rudzkis, Saulis and Statulevičius (1978), letting Φ denote the cumulative distribution function of the standard normal distribution, we deduce the Kolmogorov distance bound

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}\big(\widetilde{N}_G \leq x\big) - \Phi(x)\right| \leq \frac{C}{\lambda^{1/(4r-2)}}, \qquad \lambda \to \infty,$$

for the normalized subgraph count $\widetilde{N}_G$, see Corollary 7.1, and a moderate deviation principle, see Corollary 7.3.

In the sparse regime (20) where $c_\lambda \leq \lambda^{-\alpha}$ for some $\alpha \geq 1$, the maximal rate $\lambda^{\alpha-(\alpha-1)r}$ is attained for G a tree-like graph, and in Proposition 6.5 we obtain the cumulant bounds

$$(K_1)^r \lambda^{\alpha-(\alpha-1)r} \leq \kappa_n(N_G) \leq n!^r (K_2)^r \lambda^{\alpha-(\alpha-1)r}, \quad \lambda \geq 1,$$

if G is a tree, and

$$(K_1)^r \lambda^{r-\alpha|E(G)|} \leq \kappa_n(N_G) \leq n!^r (K_2)^r \lambda^{r-\alpha|E(G)|}, \quad \lambda \geq 1,$$

if G is a not a tree, such as e.g. a cycle graph. As a consequence of the Statulevičius condition (41), when G is a tree we find the Kolmogorov distance bound

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}\big(\widetilde{N}_G \leq x\big) - \Phi(x)\right| \leq C\lambda^{-(\alpha-(\alpha-1)r)/(4r-2)}, \qquad \lambda \to \infty,$$

provided that $1 \leq \alpha < r/(r-1)$, see Corollary 7.2.

Convergence rates in the Kolmogorov distances may be improved into classical Berry-Esseen rates when the connection function $H(x,y)$ is $\{0,1\}$-valued, e.g. in disk models as in Privault (2022), by representing subgraph counts as multiple Poisson stochastic integrals and using the fourth moment theorem for U-statistics and sums of multiple stochastic integrals Corollary 4.10 in Eichelsbacher and Thäle (2014), see also Theorem 3 in Lachièze-Rey and Reitzner (2016) or Theorem 6.3 in Privault and Serafin (2022) for Hoeffding decompositions. On the other hand, the study of stabilizing functionals

Lachièze-Rey, Schulte and Yukich (2019), Penrose and Yukich (2005) yields normal approximation with rates for random functionals represented as sums of stabilizing score functions on random geometric graphs. In the general case where $H(x,y)$ is $[0,1]$-valued, both methods no longer apply, which is why we rely on the cumulant method and the Statulevičius condition, which in turn may yield suboptimal convergence rates. Recently, moderate deviation principles have been obtained by the cumulant method for functionals of Poisson point processes in Schulte and Thäle (2023), with application to subgraph counting in random geometric graphs. However, Schulte and Thäle (2023) does not cover the random-connection model with a general connection function $H(x,y) \in [0,1]$.

This paper is organized as follows. Sections 2 and 3 introduce the preliminary framework and notations on connected partition diagrams and combinatorics of virtual cumulants that will be used for the expression of cumulants of multiparameter stochastic integrals in Section 4 and for subgraph counts in Section 5. Those expressions are applied in Section 6 to derive cumulant growth rates in the random-connection model, with application to Kolmogorov rates in subgraph counting via the Statulevičius condition in Section 7. In Section 8, normal approximation for subgraph counts on the random geometric graph is discussed under different limiting regimes.

Preliminaries

Consider a Poisson point process Ξ on $\mathbb{R}^d$, $d \geq 1$, with σ-finite intensity measure Λ on $\mathbb{R}^d$, constructed on the space

$$\Omega = \left\{\omega = \{x_i\}_{i\in I} \subset \mathbb{R}^d \ : \ \#(A \cap \omega) < \infty \text{ for all compact } A \in \mathcal{B}(\mathbb{R}^d)\right\}$$

of locally finite configurations on $\mathbb{R}^d$, whose elements $\omega \in \Omega$ are identified with the Radon point measures $\omega = \sum_{x\in\omega} \epsilon_x$, where ϵ_x denotes the Dirac measure at $x \in \mathbb{R}^d$. As in (Last and Penrose, 2018, Corollary 6.5), almost every element ω of Ω can be represented as $\omega = \{V_i\}_{1\leq i\leq N}$, where $(V_i)_{i\geq 1}$ is a random sequence in $\mathbb{R}^d$ and a $\mathbb{N} \cup \{\infty\}$-valued random variable N.

In what follows, we let $[n] := \{1,2,\dots,n\}$ for $n \geq 1$. In the next proposition, see Proposition 2 in Privault (2019), which relies on Proposition 3.1 of Privault (2012) and Lemma 2.1 of Bogdan et al. (2017), we express the moments of (1) using sums over the set $\Pi(\eta \times [r])$ of all partitions of the set

$$\eta \times [r] := \left\{(k,l) \ : \ k \in \eta,\ l = 1,\dots,r\right\}, \quad n,r \geq 1,\ \eta \subset [n].$$

Proposition 1.1. *Given $r \geq 2$, consider a connected graph G with r vertices, and a bounded measurable process of the form*

$$u(x_1,\dots,x_r) := \prod_{\{i,j\}\in E(G)} v(x_i,x_j),$$

where $v(x,y)$ is a bounded random process $v(x,y)$ independent of the underlying Poisson point process Ξ. Then, the n-th moment of the multiparameter stochastic integral

$$\sum_{\{V_1,\dots,V_r\}\subset\omega} u(V_1,\dots,V_r) = \int_{(\mathbb{R}^d)^r} u(x_1,\dots,x_r)\omega(\mathrm{d}x_1)\cdots\omega(\mathrm{d}x_r), \quad n \geq 1, \tag{1}$$

is given by the summation

$$\sum_{\rho\in\Pi([n]\times[r])} \int_{(\mathbb{R}^d)^{|\rho|}} \mathbb{E}\left[\prod_{k=1}^{n} \prod_{\{i,j\}\in E(G)} v\left(x^\rho_{k,i}, x^\rho_{k,j}\right)\right] \prod_{\eta\in\rho} \Lambda(\mathrm{d}x_\eta), \tag{2}$$

where we let $x^{\rho}_{k,l} := x_\eta$ *whenever* $(k,l) \in \eta$*, for* $\rho \in \Pi([n] \times [r])$ *and* $\eta \in \rho$*.*

2. Set partitions and diagram connectivity

Given η a finite set, we denote by $\Pi(\eta)$ the collection of its set partitions, and we let $|\sigma|$ denote the number of blocks in any partition $\sigma \in \Pi(\eta)$. Given ρ, σ two set partitions, we say that σ is coarser than ρ, or that ρ is finer than σ, and we write $\rho \le \sigma$, if every block in σ is a combination of blocks in ρ. We also denote by $\rho \vee \sigma$ the finest partition which is coarser than ρ and σ, and by $\rho \wedge \sigma$ the coarsest partition that is finer than ρ and σ. We let $\widehat{0}$ be the finest partition, which is made of a single element in each block, and we let $\widehat{1}$ be the coarsest (one-block) partition. In general, given any graph G we denote by $V(G)$ the set of its vertices, and by $E(G)$ the set of its edges.

Our study of cumulants and moments of functionals of random fields relies on partition diagrams, see Khorunzhiy (2008), Malyshev and Minlos (1991), Peccati and Taqqu (2011) and references therein for additional background. In the sequel, for $n, r \ge 1$ we let $\pi_\eta := (\pi_i)_{i\in\eta} \in \Pi(\eta \times [r])$ denote the partition made of the $|\eta|$ blocks of size r given by $\pi_k := \{(k,1), \dots, (k,r)\}$, $k \in \eta$.

Definition 2.1. Let $n, r \ge 1$. Given $\eta \subset [n]$ and $\rho \in \Pi(\eta \times [r])$ a partition of $\eta \times [r]$, we denote by $\Gamma(\rho, \pi_\eta)$ the diagram, or graphical representation of the partition ρ, constructed by:

1. arranging the elements of $\eta \times [r]$ into an array of $|\eta|$ rows and r columns, and
2. connecting all elements within a same block of ρ by a tree graph.

In addition, we say that the partition diagram $\Gamma(\rho, \pi_\eta)$ is connected when $\rho \vee \pi_\eta = \widehat{1}$.

For shortness of notation, in the sequel we say that a partition ρ is connected when its diagram $\Gamma(\rho, \pi)$ is connected. For example, taking $\eta := \{2,3,5,8,10\}$, given the partitions

$$\begin{aligned}\rho = \big\{&\{(2,1),(3,1),(3,2),(3,3)\},\{(2,2),(2,3),(2,4),(3,4)\},\{(5,1)\},\{(5,2),(8,2)\},\\ &\{(5,3)\},\{(5,4),(8,3)\},\{(8,1),(10,1)\},\{(8,4)\},\{(10,2),(10,3),(10,4)\}\big\}\end{aligned}$$

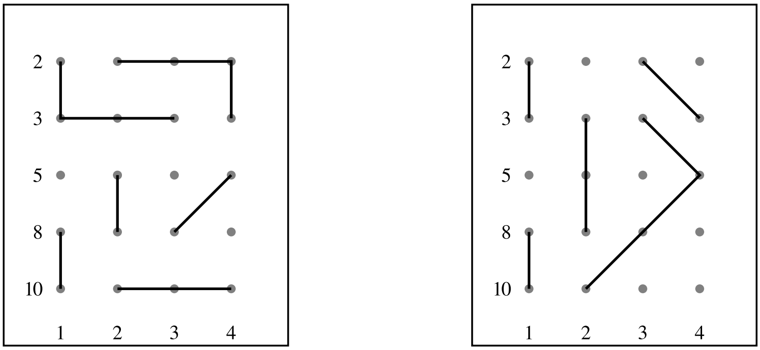
and

$$\begin{aligned}\sigma = \big\{&\{(2,1),(3,1)\},\{(2,2)\},\{(2,3),(3,4)\},\{(2,4)\},\{(3,2),(5,2),(8,2)\},\\ &\{(3,3),(5,4),(8,3),(10,2)\},\{(5,1)\},\{(5,3)\},\{(8,1),(10,1)\},\{(8,4)\},\{(10,3)\},\{(10,4)\}\big\},\end{aligned}$$

of $\eta \times [4]$, Figure 1-*a*) presents an example of a non-connected partition diagram $\Gamma(\rho, \pi)$, and Figure 1-*b*) presents an example of a connected partition diagram $\Gamma(\sigma, \pi)$.
Note that the above notion of connected partition diagram is distinct from that of irreducible partition, see, e.g., Bender, Odlyzko and Richmond (1985).

Definition 2.2. Let $n \ge 1$, G a connected graph with $|V(G)| = r$ vertices, $r \ge 1$, and consider $G_1, \dots, G_n$ copies of G respectively built on $\pi_1, \dots, \pi_n$. Let also $\rho \in \Pi([n] \times [r])$ be a partition of $[n] \times [r]$.

1. We let $\widetilde{\rho}_G$ be the multigraph constructed on the blocks of ρ by adding an edge between two blocks ρ_1, ρ_2 of the partition ρ whenever there exist $(k, l_1) \in \rho_1$ and $(k, l_2) \in \rho_2$ such that (l_1, l_2) is an edge in G_k.
2. We let ρ_G be the graph constructed on the blocks of ρ by removing redundant edges in $\widetilde{\rho}_G$, so that at most one edge remains between any two blocks $\rho_1, \rho_2 \in \rho$.



(a) Non-connected partition diagram $\Gamma(\rho, \pi)$. (b) Connected partition diagram $\Gamma(\sigma, \pi)$.

Figure 1. Two examples of partition diagrams with $\eta = \{2,3,5,8,10\}$, $n = 10$, $r = 4$.

Figure 2-b) presents an illustration of the multigraph $\widetilde{\rho}_G$ and graph ρ_G on the blocks of ρ when G is the line graph $\{(1,2),(2,4),(3,4)\}$ on $\{1,2,3,4\}$.

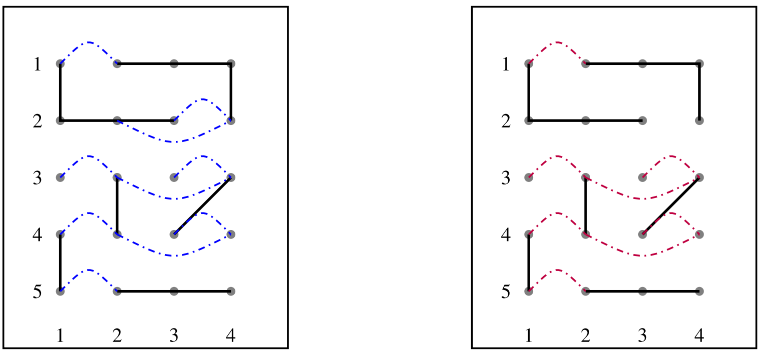
Definition 2.3. Let $n, r \geq 1$, and let $\rho \in \Pi([n] \times [r])$ be a partition of $[n] \times [r]$.

1. For $b \subset [n]$, we let $\rho_b \subset \rho$ be defined as

$$\rho_b := \{c \in \rho \ : \ c \subset b \times [r]\}.$$

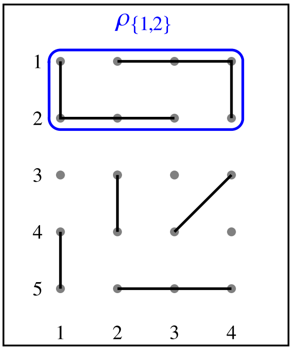
2. Given $\eta \subset [n]$ we split any partition ρ of $\eta \times [r]$ into the equivalence classes deduced from the connected components of the graph ρ_G, as

$$\rho = \bigcup_{\substack{b \subset \eta \\ b \times [r] \in \rho \vee \pi_\eta}} \rho_b. \tag{3}$$

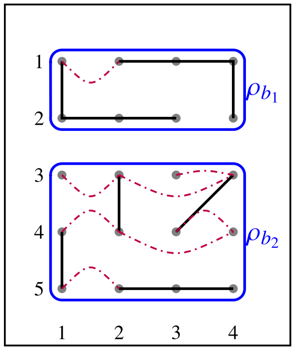


(a) Diagram $\Gamma(\rho, \pi)$ and multigraph $\widetilde{\rho}_G$ in blue. (b) Diagram $\Gamma(\rho, \pi)$ and graph ρ_G in red.

Figure 2. Diagram and graphs G, ρ_G, $\widetilde{\rho}_G$ with $n = 5$, $r = 4$.



(a) Diagram $\Gamma(\rho,\pi)$ and block $\rho_{\{1,2\}}$.



(b) Splitting $\{\rho_{b_1}, \rho_{b_2}\}$ of ρ according to ρ_G.

Figure 3. Splitting of the partition ρ with $\rho \vee \pi = \{\pi_1 \cup \pi_2, \pi_3 \cup \pi_4 \cup \pi_5\}$ and $n = 5$, $r = 4$.

As an example, in Figure 3-a), when $b = \{1,2\}$ we have

$$\rho_{\{1,2\}} = \big\{\{(1,1),(2,1),(2,2),(2,3)\},\{(1,2),(1,3),(1,4),(2,4)\}\big\},$$

and the partition (3) is illustrated in Figure 3-b) with $b_1 = \{1,2\}$ and $b_2 = \{3,4,5\}$.

Definition 2.4. Let $n,r \geq 1$. Given $\sigma \in \Pi([n])$ a partition of $[n]$, we let $\Pi_\sigma([n]\times[r])$ denote the set of partitions ρ of $[n]\times[r]$ such that

$$\rho \vee \pi = \{b \times [r] \ : \ b \in \sigma\},$$

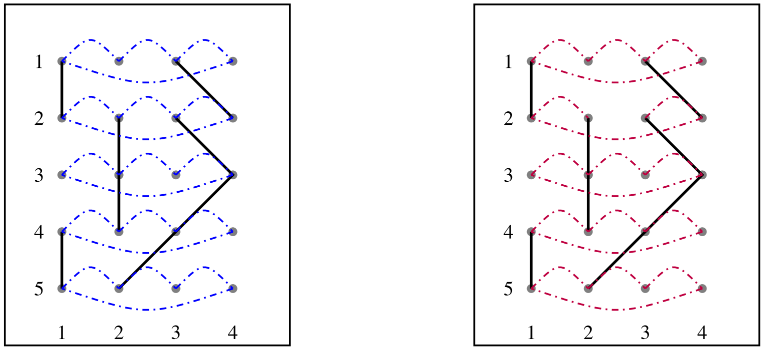
and we partition $\Pi([n]\times[r])$ as

$$\Pi([n]\times[r]) = \bigcup_{\sigma\in\Pi([n])} \Pi_\sigma([n]\times[r]). \tag{4}$$

We note that given $\eta \subset [n]$, the set $\Pi_{\widehat{1}}(\eta\times[r])$ consists of the partitions ρ of $\eta\times[r]$ for which the graph ρ_G is connected, as in Figure 4.
The following lemma will be useful when applying induction arguments on connected set partitions in $\Pi_{\widehat{1}}([n]\times[r])$.

Lemma 2.5. *Let $n \geq 2$. For any connected partition $\rho \in \Pi_{\widehat{1}}([n]\times[r])$ there exists $i \in \{1,\ldots,n\}$ such that the set partition $\{b\backslash\pi_i : b \in \rho\}$ of $\{1,\ldots,i-1,i+1,\ldots,n\}\times[r]$ is connected.*

Proof. Let $\rho \in \Pi_{\widehat{1}}([n]\times[r])$. We consider the connected undirected graph $\mathfrak{g}$ on $[n]$ in which two vertices $i,j \in [n]$ are connected if and only if there exists a block $b \in \rho$ such that $\pi_i \cap b \neq \varnothing$ and $\pi_j \cap b \neq \varnothing$, see Figure 5-$a$) for an example with $n = 5$.
By e.g. Theorem 4.2.3 in Balakrishnan and Ranganathan (2012), $\mathfrak{g}$ contains a spanning tree $\bar{\mathfrak{g}}$, as shown in Figure 5-b). Let $i \in [n]$ be a leaf in the tree $\bar{\mathfrak{g}}$. If the partition $\rho^{(i)} := \{b \setminus \pi_i \ : \ b \in \rho\}$ of $([n]\setminus\{i\})\times[r]$ had more than one connected component, then, for ρ to be connected, π_i would have to connect to all such components, hence the vertex i would be adjacent to more than one vertex in $\bar{\mathfrak{g}}$, which is not the case. $\square$



(a) Diagram $\Gamma(\rho,\pi)$ and multigraph $\widetilde{\rho}_G$ in blue. (b) Diagram $\Gamma(\rho,\pi)$ and graph ρ_G in red.

Figure 4. Connected non-flat partition diagram with G a cycle graph and $n = 5$, $r = 4$.

In what follows, we say that a partition ρ of $[n]\times[r]$ is *non-flat* if its diagram $\Gamma(\rho,\pi)$ is non-flat, i.e. if $\rho \wedge \pi = \widehat{0}$, see Chapter 4 of Peccati and Taqqu (2011) and Figure 4.

Definition 2.6. Given $n, r \geq 1$, we denote by

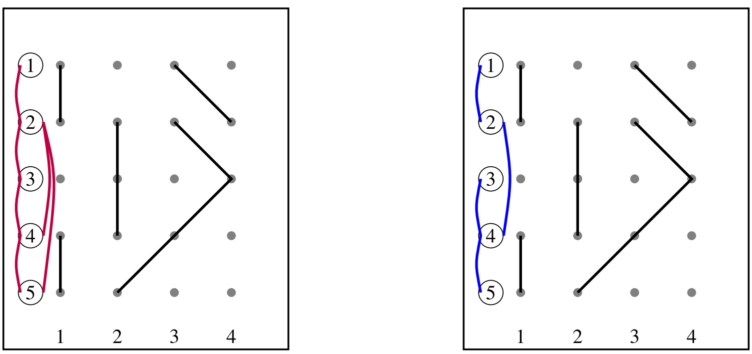
$$\mathrm{NF}(n,r) := \{\rho \in \Pi([n]\times[r]) \ : \ \rho \wedge \pi = \widehat{0}\}$$

the set of non-flat partitions of $[n]\times[r]$, and by

$$\mathrm{CNF}(n,r) := \{\rho \in \Pi_{\widehat{1}}([n]\times[r]) \ : \ \rho \wedge \pi = \widehat{0}\}$$

the set of connected non-flat partitions of $[n]\times[r]$.

We will also consider the following set of connected non-flat partitions which have a maximal number of blocks.



(a) Diagram $\Gamma(\rho,\pi)$ and graph $\mathfrak{g}$. (b) Diagram $\Gamma(\rho,\pi)$ and spanning tree $\bar{\mathfrak{g}}$.

Figure 5. Example of graph $\mathfrak{g}$ and its spanning tree subgraph.

Definition 2.7. Given $n \geq 1$ and $r \geq 2$, we denote by

$$\mathrm{M}(n,r) := \{\rho \in \mathrm{CNF}(n,r) \ : \ |\rho| = 1 + (r-1)n\}$$

the set of maximal connected non-flat partitions of $[n] \times [r]$.

The bound in part (a) of the next lemma is consistent with (6.2) in Proposition 6.1 of Schulte and Thäle (2023), which shows that the power r of $n!$ cannot be improved in (5).

Lemma 2.8. *a) The cardinality of the set* $\mathrm{NF}(n,r)$ *of non-flat partitions of* $[n] \times [r]$ *satisfies*

$$|\mathrm{NF}(n,r)| \leq n!^r r!^{n-1}, \qquad n,r \geq 1. \tag{5}$$

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

$$|\mathrm{M}(n,r)| = r^{n-1} \prod_{i=1}^{n-1} (1 + (r-1)i), \qquad n,r \geq 1, \tag{6}$$

with the bounds

$$((r-1)r)^{n-1}(n-1)! \leq |\mathrm{M}(n,r)| \leq ((r-1)r)^{n-1} n!, \quad n \geq 1, \ r \geq 2. \tag{7}$$

Proof. a) We clearly have $|\mathrm{NF}(1,r)| = 1$ for all $r \geq 1$. Any non-flat partition $\rho \in \mathrm{NF}(n+1,r)$ can be obtained from a non-flat partition in $\mathrm{NF}(n,r)$ in at most $(n+1)^r r!$ ways, by connecting each of the r new points in at most $n+1$ possible ways (including non-connection), and multiplying by the number $r!$ of possible permutations of π_i. This yields the induction inequality

$$|\mathrm{NF}(n+1,r)| \leq (n+1)^r r! |\mathrm{NF}(n,r)|,$$

from which we conclude that (5) holds.
b) We clearly have $|\mathrm{M}(1,r)| = 1$ for all $r \geq 1$. Next, each maximal connected non-flat partition $\rho \in \mathrm{M}(n+1,r)$ can be obtained by choosing one of r elements of $\{(n+1,1), \ldots, (n+1,r)\}$, and connecting them in $1 + (r-1)n$ ways to any partition in $\mathrm{M}(n,r)$, $n \geq 1$. This implies the recursion formula

$$|\mathrm{M}(n+1,r)| = r \times (1 + (r-1)n) |\mathrm{M}(n,r)|,$$

which yields (6). □

3. Virtual cumulants

The following definition uses the concept of independence of a virtual field with respect to graph connectedness, see Relation (17) in (Malyshev and Minlos, 1991, p. 34).

Definition 3.1. Let $n,r \geq 1$. We say that a mapping F defined on partitions of $[n] \times [r]$ admits the *connectedness factorization* property if it decomposes according to the partition (3) as

$$F(\rho) = \prod_{b \times [r] \in \rho \vee \pi} F(\rho_b), \qquad \rho \in \Pi([n] \times [r]). \tag{8}$$

In what follows, given F a mapping defined on the partitions of $[n] \times [r]$, we will use the Möbius transform $\widehat{F}$ of F, defined as

$$\widehat{F}(\eta) := \sum_{\rho \in \Pi(\eta \times [r])} F(\rho), \qquad \eta \subset [n],$$

with $\widehat{F}(\varnothing) := 0$, see Rota (1964) and § 2.5 of Peccati and Taqqu (2011). We refer to (Malyshev and Minlos, 1991, p. 33) for the following definition.

Definition 3.2. Let $n, r \geq 1$. The virtual cumulant G of a mapping F on $\bigcup_{\eta \subset [n]} \Pi(\eta \times [r])$ is defined by letting $C_F(\eta) := \widehat{F}(\eta)$ when $|\eta| = 1$, and then recursively by

$$C_F(\eta) := \widehat{F}(\eta) - \sum_{\substack{\sigma \in \Pi(\eta) \\ |\sigma| \geq 2}} \prod_{b \in \sigma} C_F(b), \qquad \eta \subset [n], \quad |\eta| \geq 2. \tag{9}$$

Relation (9) also implies the relation

$$C_F(\eta) = \sum_{\sigma \in \Pi(\eta)} (-1)^{|\sigma|-1} (|\sigma| - 1)! \prod_{b \in \sigma} \widehat{F}(b), \tag{10}$$

see Relation (16') page 33 of Malyshev and Minlos (1991), which is also the classical cumulant-moment relationship, see e.g. Corollary 3.2.2 in Peccati and Taqqu (2011). The following proposition is an extension of the classical Lemma 2 in (Malyshev and Minlos, 1991, p. 34), see also Lemma 3.1 in Khorunzhiy (2008).

Proposition 3.3. *Let $n, r \geq 1$. Let F be a mapping defined on $\bigcup_{\eta \subset [n]} \Pi(\eta \times [r])$ and admitting the connectedness factorization property (8). Then, for $\eta \subset [n]$ with $\eta \neq \varnothing$, the virtual cumulant of F is given by the sum*

$$C_F(\eta) = \sum_{\substack{\sigma \in \Pi_{\widehat{1}}(\eta \times [r]) \\ \text{(connected)}}} F(\sigma) \tag{11}$$

over connected partitions on $\eta \times [r]$.

Proof. The claim is true when $|\eta| = 1$. Assume that it is true for all $\eta \subset [n]$ for some $n \geq 1$, and let η be such that $|\eta| = n + 1$. By (4) and (8), we have

$$\begin{aligned}
\widehat{F}(\eta) &= \sum_{\rho \in \Pi(\eta \times [r])} F(\rho) \\
&= \sum_{\sigma \in \Pi(\eta)} \sum_{\rho \in \Pi_\sigma(\eta \times [r])} F(\rho) \\
&= \sum_{\sigma \in \Pi(\eta)} \sum_{\rho \in \Pi_\sigma(\eta \times [r])} \prod_{b \in \sigma} F(\rho_b) \\
&= \sum_{\sigma \in \Pi(\eta)} \prod_{b \in \sigma} \sum_{\substack{\rho \in \Pi_{\widehat{1}}(b \times [r]) \\ \text{(connected)}}} F(\rho)
\end{aligned}$$

$$= \sum_{\substack{\rho \in \Pi_{\widehat{1}}(\eta \times [r]) \\ \text{(connected)}}} F(\rho) + \sum_{\substack{\sigma \in \Pi(\eta) \\ |\sigma| \geq 2}} \prod_{b \in \sigma} C_F(b),$$

where the last equality follows from the induction hypothesis (11) when $|\eta| \leq n$. The proof is completed by subtracting the last term on both sides. $\square$

4. Cumulants of multiparameter stochastic integrals

Proposition 4.1 rewrites the product in (2) of Proposition 1.1 as a product on the edges of the graph ρ_G similarly to Proposition 4 in Privault (2019) when $v(x,y)$ vanishes on the diagonal, and it generalizes Proposition 2.4 of Jansen (2019) from two-parameter Poisson stochastic integrals to multiparameter integrals of higher orders.

Proposition 4.1. *Let $n \geq 1$, $r \geq 2$, and assume that the process $v(x,y)$ vanishes on diagonals, i.e. $v(x,x) = 0$, $x \in \mathbb{R}^d$. Then, the n-th moment of the multiparameter stochastic integral (1) is given by the summation*

$$\sum_{\substack{\rho \in \Pi([n]\times[r]) \\ \rho \wedge \pi = \widehat{0} \\ (non-flat)}} \int_{(\mathbb{R}^d)^{|\rho|}} \prod_{\{\eta_1,\eta_2\} \in E(\rho_G)} \mathbb{E}\big[v(x_{\eta_1}, x_{\eta_2})^{m(\eta_1,\eta_2)}\big] \prod_{\eta \in V(\rho_G)} \Lambda(\mathrm{d}x_\eta),$$

over connected non-flat partitions, where $m(\eta_1,\eta_2)$ represents the multiplicity of the edge (η_1,η_2) in the multigraph $\widetilde{\rho}_G$.

The next proposition is a consequence of Propositions 3.3 and 4.1, and it also extends Proposition 2.5 of Jansen (2019) from the two-parameter case to the multiparameter case. Note that in our setting, the two-parameter case only applies to the edge counting.

Proposition 4.2. *Let $n \geq 1$, $r \geq 2$, and assume that the process $v(x,y)$ vanishes on diagonals, i.e. $v(x,x) = 0$, $x \in \mathbb{R}^d$. Then, the n-th cumulant of the multiparameter stochastic integral (1) is given by the summation*

$$\sum_{\substack{\rho \in \Pi_{\widehat{1}}([n]\times[r]) \\ \rho \wedge \pi = \widehat{0} \\ \text{(non-flat connected)}}} \int_{(\mathbb{R}^d)^{|\rho|}} \prod_{\{\eta_1,\eta_2\} \in E(\rho_G)} \mathbb{E}\big[v(x_{\eta_1}, x_{\eta_2})^{m(\eta_1,\eta_2)}\big] \prod_{\eta \in V(\rho_G)} \Lambda(\mathrm{d}x_\eta) \tag{12}$$

over connected non-flat partitions.

Proof. The functional

$$F(\rho) := \int_{(\mathbb{R}^d)^{|\rho|}} \prod_{\{\eta_1,\eta_2\} \in E(\rho_G)} \mathbb{E}\big[v(x_{\eta_1}, x_{\eta_2})^{m(\eta_1,\eta_2)}\big] \prod_{\eta \in V(\rho_G)} \Lambda(\mathrm{d}x_\eta)$$

satisfies the connectedness factorization property (8), as for $\sigma = b \times [r] \in \rho \vee \pi$ and $\sigma' = b' \times [r] \in \rho \vee \pi$ with $b \neq b'$, the variables $(x_\eta)_{\eta \in \rho_b}$ are distinct from the variables $(x_\eta)_{\eta \in \rho_{b'}}$ in the above integration. Hence, Relation (12) follows from Proposition 3.3 and the classical cumulant-moment relationship (10), since by Proposition 1.1, $\widehat{F}([n])$ is the n-th moment of the multiparameter stochastic integral (1). $\square$

5. Cumulants of subgraph counts

Let $H : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ denote a measurable connection function such that

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

for all $y \in \mathbb{R}$. Given $\omega \in \Omega$, for any $x, y \in \omega$ with $x \neq y$, an edge connecting x and y is added with probability $H(x,y)$, independently of the other pairs. The resulting random graph, together with the Poisson point process Ξ with intensity measure Λ on $\mathbb{R}^d$, is called the random-connection model and denoted by $G_H(\Xi)$.

In the case where the connection function H is given by $H(x,y) := \mathbf{1}_{\{\|x-y\| \leq R\}}$ for some $R > 0$, the resulting graph is completely determined by the geometry of the underlying point process Ξ, and is called a random geometric graph Penrose (2003), which is included as a special case in this paper. The main case of interest in this section is the general random-connection model. We shall have further discussion about the random geometric graph in Section 8, as we believe it is of independent interest.

Given G a connected graph with $|V(G)| = r$ vertices, we denote N_G the count of subgraphs isomorphic to G in the random-connection model $G_H(\Xi)$, which can be represented as the multiparameter stochastic integral

$$N_G := \sum_{\{V_1,\ldots,V_r\} \subset \omega} \prod_{\{i,j\} \in E(G)} \mathbf{1}_{\{V_i \leftrightarrow V_j\}} = \int_{(\mathbb{R}^d)^r} \prod_{\{i,j\} \in E(G)} \mathbf{1}_{\{x_i \leftrightarrow x_j\}} \omega(\mathrm{d}x_1) \cdots \omega(\mathrm{d}x_r),$$

up to division by the number of automorphisms of G. Here, $\mathbf{1}_{\{x \leftrightarrow y\}}$ denotes a $\{0,1\}$-valued Bernoulli random variable with parameter $H(x,y)$ when $x \neq y$, $x, y \in \mathbb{R}^d$, and

$$\mathbf{1}_{\{x \leftrightarrow x\}} := 0, \qquad x \in \mathbb{R}^d. \tag{13}$$

Consequently, we have $\mathbf{1}_{\{V_i \leftrightarrow V_j\}} = 1$ or 0 depending on whether V_i and V_j are connected or not by an edge in $G_H(\Xi)$. The first moment of N_G can be computed as

$$\mathbb{E}[N_G] = \int_{(\mathbb{R}^d)^r} \left(\prod_{\{i,j\} \in E(G)} H(x_i, x_j) \right) \prod_{i=1}^{r} \Lambda(\mathrm{d}x_i). \tag{14}$$

Higher order moments can be computed from the following result which is a direct consequence of Proposition 4.2 by taking $v(x,y) := \mathbf{1}_{\{x \leftrightarrow y\}}$ in (12) and by using *non-flat* partition diagrams $\Gamma(\rho, \pi)$ such that $\rho \wedge \pi = \widehat{0}$, to take into account condition (13).

Proposition 5.1. *Let $n \geq 1$ and $r \geq 2$. The moments and cumulants of N_G are given by the summation*

$$\mathbb{E}[(N_G)^n] = \sum_{\substack{\rho \in \Pi([n] \times [r]) \\ \rho \wedge \pi = \widehat{0} \\ (\text{non-flat})}} \int_{(\mathbb{R}^d)^{|\rho|}} \left(\prod_{\{\eta_1, \eta_2\} \in E(\rho_G)} H(x_{\eta_1}, x_{\eta_2}) \right) \prod_{\eta \in V(\rho_G)} \Lambda(\mathrm{d}x_\eta), \tag{15}$$

over non-flat partitions, and by the summation

$$\kappa_n(N_G) = \sum_{\substack{\rho \in \Pi_{\widehat{1}}([n] \times [r]) \\ \rho \wedge \pi = \widehat{0} \\ (\text{non-flat connected})}} \int_{(\mathbb{R}^d)^{|\rho|}} \left(\prod_{\{\eta_1, \eta_2\} \in E(\rho_G)} H(x_{\eta_1}, x_{\eta_2}) \right) \prod_{\eta \in V(\rho_G)} \Lambda(\mathrm{d}x_\eta), \tag{16}$$

over connected non-flat partitions.

Proof. Relations (15)-(16) are consequence of Proposition 4.2, after taking $v(x_i,x_j) := \mathbf{1}_{\{x_i \leftrightarrow x_j\}}$, $\{i,j\} \in E(G)$. The summations are restricted to *non-flat* partitions due to condition (13) as in Section 2 of Privault (2019). □

6. Asymptotic growth of subgraph count cumulants

In this section we consider the following assumption, where $(\Lambda_\lambda)_{\lambda>0}$ is a family of σ-finite intensity measures on $\mathbb{R}^d$ and $H(x,y)$ is the connection function of the random-connection model.

Assumption 6.1. Let $r \geq 2$ and $n \geq 1$. There exist constants $c_H, C_H > 0$ such that for any connected non-flat partition $\rho \in \Pi_{\widehat{1}}([n] \times [r])$, we have

$$c_H^{|E(\rho_G)|}(\lambda C_H)^{|V(\rho_G)|} \leq \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \left(\prod_{\{i,j\} \in E(\rho_G)} H(x_i,x_j) \right) \prod_{k \in V(\rho_G)} \Lambda_\lambda(\mathrm{d}x_k), \quad \lambda > 0. \tag{17}$$

We consider two settings satisfying Assumption 6.1.

Example (Increasing intensity). When the intensity measure Λ_λ takes the form

$$\Lambda_\lambda(\mathrm{d}x) = \lambda \mu(\mathrm{d}x), \qquad \lambda > 0,$$

for μ a finite diffuse measure on $\mathbb{R}^d$, Assumption 6.1 is satisfied by any translation-invariant continuous kernel function $H : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ non vanishing at $(0,0)$. Indeed, in this case there exist $c_H > 0$ and a Borel set $B \subset \mathbb{R}^d$ such that $\mu(B) > 0$ and

$$H(x,y) = H(x-y,0) \geq c_H \mathbf{1}_B(x) \mathbf{1}_B(y), \qquad x,y \in \mathbb{R}^d,$$

hence

$$\begin{aligned} c_H^{|E(\rho_G)|}(\mu(B))^{|V(\rho_G)|} &= c_H^{|E(\rho_G)|} \int_B \cdots \int_B \prod_{k \in V(\rho_G)} \mu(\mathrm{d}x_k) \\ &\leq \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \left(\prod_{\{i,j\} \in E(\rho_G)} H(x_i,x_j) \right) \prod_{k \in V(\rho_G)} \mu(\mathrm{d}x_k), \end{aligned}$$

so that we can take $C_H := \mu(B)$.

The setting of the above increasing intensity example covers the following long and short range dependence settings:

1. the power-law fading kernel $H(x,y) = 1 \wedge \|x-y\|^{-\beta}$, $x,y \in \mathbb{R}^d$, for some $\beta > 0$,
2. the Rayleigh fading kernel $H(x,y) = e^{-\beta\|x-y\|^2}$, $x,y \in \mathbb{R}^d$, for some $\beta > 0$,
3. the Boolean kernel $H(x,y) = \mathbf{1}_{\{\|x-y\| \leq R\}}$, $x,y \in \mathbb{R}^d$, for some $R > 0$, which yields the random geometric graph, with $C_H = v_d(R/2)^d$, where v_d denotes the volume of the unit ball in $\mathbb{R}^d$.

Example (Growing observation window). When the intensity measure Λ_λ takes the form

$$\Lambda_\lambda(\mathrm{d}x) = \mathbf{1}_{A_\lambda}(x)\mu(\mathrm{d}x), \qquad \lambda > 0,$$

where $(A_\lambda)_{\lambda>0}$ is a non-decreasing sequence of Borel subsets of $\mathbb{R}^d$ such that $A_\lambda \uparrow \mathbb{R}^d$ as λ tends to infinity, μ a diffuse σ-finite measure on $\mathbb{R}^d$, and the kernel function $H : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ is lower bounded as $H(x,y) \geq c_H$, $x,y \in \mathbb{R}^d$ for some $c_H > 0$, Assumption 6.1 is satisfied with $\mu(A_\lambda) = C_H\lambda$, e.g. when A_λ is the ball of radius $\lambda^{1/d} > 0$ and μ is the Lebesgue measure on $\mathbb{R}^d$, with $C_H = v_d$.

Next, we investigate the asymptotic behaviour of the cumulants $\kappa_n(N_G)$ as the intensity λ tends to infinity, as a consequence of the partition diagram representation of cumulants. In what follows, given two positive functions f and g on $(1,\infty)$ we write $f(\lambda) \ll g(\lambda)$ if $\lim_{\lambda\to\infty} g(\lambda)/f(\lambda) = \infty$.

Definition 6.2. Let G be a connected graph with $|V(G)| = r$ vertices, $r \geq 2$. For every $\lambda > 0$, let $G_{H_\lambda}(\Xi)$ denote the random-connection model with connection function

$$H_\lambda(x,y) := c_\lambda H(x,y), \qquad x,y \in \mathbb{R}^d.$$

We consider the following regimes.

- Dilute regime: for some constant $K > 0$ we have

$$\lambda^{-1/\zeta} \ll c_\lambda \leq K, \qquad \lambda \to \infty, \tag{18}$$

  where

$$\zeta := \max\left\{ \frac{|E(H)|}{|V(H)| - 1} \ : H \subseteq G, |V(H)| \geq 2 \right\}. \tag{19}$$

- Sparse regime: for some constants $K > 0$ and $\alpha \geq 1$ we have

$$c_\lambda \leq \frac{K}{\lambda^\alpha}, \qquad \lambda \to \infty. \tag{20}$$

In case $c_\lambda = K$ for all $\lambda > 0$ we also say that we are in the full random graph regime, and in the sequel we take $K = 1$ for simplicity. We note that in general we have $\zeta > 1$ except when G is a tree, in which case $\zeta = 1$.

Proposition 6.3 (Dilute regime). *Let G be a connected graph with $|V(G)| = r$ vertices, $r \geq 2$, satisfying Assumption 6.1 for all $n \geq 1$ in the dilute regime (18). We have the cumulant bounds*

$$(n-1)! c_\lambda^{n|E(G)|} (K_1\lambda)^{1+(r-1)n} \leq \kappa_n(N_G) \leq n!^r c_\lambda^{n|E(G)|} (K_2\lambda)^{1+(r-1)n}, \quad \lambda \geq 1, \tag{21}$$

for some constants K_1, $K_2 > 0$ independent of λ and $n \geq 1$.

Proof. We identify the leading terms in the sum (16) over connected non-flat partitions, in which every summand involves a factor $c_\lambda^{|E(\rho_G)|}\lambda^{|V(\rho_G)|}$, since every vertex in ρ_G contributes a factor λ, and that every edge contributes a factor c_λ. Therefore, it is sufficient to show that for any ρ_G with $\rho \in \mathrm{CNF}(n,r)$ a connected non-flat partition of $[n] \times [r]$, we have

$$c_\lambda^{|E(\rho_G)|}\lambda^{|V(\rho_G)|} = O\big(\lambda^{1+(r-1)n} c_\lambda^{n|E(G)|}\big). \tag{22}$$

This claim follows from (14) when $n = 1$. Suppose that (22) holds up to the rank $n \geq 1$, and let $\rho \in \Pi_{\widehat{1}}([n+1]\times[r])$ be a connected non-flat partition. By Lemma 2.5, there exists $i \in [n+1]$ such that the subgraph $\bar{\rho}_G$ induced by ρ_G on the vertex set

$$V(\bar{\rho}_G) := \left\{ b \in \rho \,:\, b \cap (\cup_{j\neq i}\pi_j) \neq \varnothing \right\}$$

is connected. Let $\widehat{\rho}_G$ denote the subgraph induced by ρ_G on the vertex set

$$V(\widehat{\rho}_G) := \left\{ b \in \rho \,:\, b \cap \pi_i \neq \varnothing \right\},$$

with $\widehat{\rho}_G \simeq G$ because ρ is non-flat, and let $H := \bar{\rho}_G \cap \widehat{\rho}_G$. Since $H \subseteq \widehat{\rho}_G$ we have

$$\lambda^{|V(H)|-1} c_\lambda^{|E(H)|} \geq \left(\lambda c_\lambda^{\zeta}\right)^{|E(H)|/\zeta}, \qquad \lambda \geq 1.$$

Hence from $\lim_{\lambda\to\infty} \lambda c_\lambda^{\zeta} = \infty$ we get

$$\liminf_{\lambda\to\infty} \lambda^{|V(H)|-1} c_\lambda^{|E(H)|} > 0.$$

On the other hand, by the induction hypothesis we have

$$\frac{\lambda^{|V(\bar{\rho}_G)|} c_\lambda^{|E(\bar{\rho}_G)|}}{\lambda^{1+(r-1)n} c_\lambda^{n|E(G)|}} = O(1), \tag{23}$$

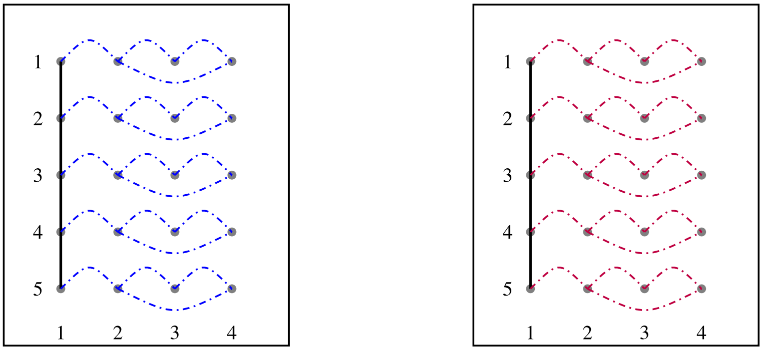
hence, since $|V(\widehat{\rho}_G)| = |V(G)|$ and $|E(\widehat{\rho}_G)| = |E(G)|$,

$$\begin{aligned}
\frac{\lambda^{|\rho|} c_\lambda^{|E(\rho_G)|}}{\lambda^{1+(r-1)(n+1)} c_\lambda^{(n+1)|E(G)|}} &= \frac{\lambda^{|V(\bar{\rho}_G)|+|V(\widehat{\rho}_G)|-|V(H)|} c_\lambda^{|E(\bar{\rho}_G)|+|E(\widehat{\rho}_G)|-|E(H)|}}{\lambda^{1+(r-1)(n+1)} c_\lambda^{(n+1)|E(G)|}} \\
&= \frac{\lambda^{|V(\bar{\rho}_G)|} c_\lambda^{|E(\bar{\rho}_G)|}}{\lambda^{1+(r-1)n} c_\lambda^{n|E(G)|}} \cdot \frac{\lambda^{|V(\widehat{\rho}_G)|} c_\lambda^{|E(\widehat{\rho}_G)|}}{\lambda^{r} c_\lambda^{|E(G)|}} \cdot \frac{\lambda^{-|V(H)|} c_\lambda^{-|E(H)|}}{\lambda^{-1}} \\
&= \frac{\lambda^{|V(\bar{\rho}_G)|} c_\lambda^{|E(\bar{\rho}_G)|}}{\lambda^{1+(r-1)n} c_\lambda^{n|E(G)|}} \cdot \left(\lambda^{|V(H)|-1} c_\lambda^{|E(H)|}\right)^{-1} \\
&= O(1),
\end{aligned}$$

therefore (22) holds at the rank $n+1$. As a consequence, the leading terms in (16) are those associated with the connected partition diagrams $\Gamma(\rho,\pi)$ having the highest block count, i.e. which have $1+(r-1)n$ blocks, see Figure 6 for a sample of such a partition diagram.

Finally, we observe that any maximal connected non-flat partition $\rho \in \mathrm{M}(n,r)$ satisfies $|E(\rho_G)| = n \times |E(G)|$, as can be checked in Figure 6. Therefore, by (5)-(7), (16) and (17), we obtain

$$\begin{aligned}
&c^{n|E(G)|} C^{1+(r-1)n} c_\lambda^{n|E(G)|} ((r-1)r)^{n-1} (n-1)! \lambda^{1+(r-1)n} \\
&\leq \lambda^{1+(r-1)n} \sum_{\rho\in \mathrm{M}(n,r)} \int_{(\mathbb{R}^d)^{1+(r-1)n}} \left(\prod_{\{\eta_1,\eta_2\}\in E(\rho_G)} H_\lambda(x_{\eta_1}, x_{\eta_2}) \right) \prod_{\eta\in V(\rho_G)} \mu(\mathrm{d}x_\eta), \\
&\leq \kappa_n(N_G) \\
&\leq n!^r r!^{n-1} (1+\mu(\mathbb{R}^d))^{1+(r-1)n} c_\lambda^{n|E(G)|} \lambda^{1+(r-1)n},
\end{aligned}$$



(a) Diagram $\Gamma(\rho, \pi)$ and graph $\widetilde{\rho}_G$ in blue. (b) Diagram $\Gamma(\rho, \pi)$ and graph ρ_G in red.

Figure 6. Example of maximal connected partition diagram with $n = 5$ and $r = 4$.

which yields (21). □

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

$$\widetilde{N}_G := \frac{N_G - \kappa_1(N_G)}{\sqrt{\kappa_2(N_G)}}, \qquad n \geq 1.$$

The following result shows that for $n \geq 3$ the normalized cumulant $\kappa_n(\widetilde{N}_G)$ tends to zero in (24), hence $\widetilde{N}_G$ converges in distribution to the normal distribution by Theorem 1 in Janson (1988).

Corollary 6.4 (Dilute regime). *Let G be a connected graph with $|V(G)| = r$ vertices, $r \geq 2$, satisfying Assumption 6.1 for $n = 2$ in the dilute regime (18). We have the normalized cumulant bounds*

$$\left|\kappa_n\left(\widetilde{N}_G\right)\right| \leq n!^r (K\lambda)^{-(n/2-1)}, \qquad \lambda \geq 1, \quad n \geq 2, \tag{24}$$

where $K > 0$ is a constant independent of $\lambda > 0$ and $n \geq 1$.

Proof. We note that the upper bound in (21) does not require Assumption 6.1, hence we have, for $n \geq 2$,

$$\left|\kappa_n(\widetilde{N}_G)\right| \leq \frac{n!^r c_\lambda^{n|E(G)|}(K_2\lambda)^{1+(r-1)n}}{\left((2-1)! c_\lambda^{2|E(G)|}(K_1\lambda)^{1+2(r-1)}\right)^{n/2}} = K_2 \left(\frac{(K_2/K_1)^{r-1}}{\sqrt{K_1}}\right)^n n!^r \lambda^{-(n/2-1)}.$$

□

The following result yields a positive cumulant growth of order $\alpha - (\alpha - 1)r > 0$ in (25) for trees in the sparse regime with $\alpha \in [1, r/(r-1))$, while in the case of non-tree graphs such as cycle graphs the growth rate exponent $r - \alpha|E(G)| \leq (1-\alpha)r \leq 0$ is negative or zero in (26) and (27). In addition, the normalized cumulant $\kappa_n(\widetilde{N}_G)$ tends to zero for $n \geq 3$ in (29) only when G is a tree, in which case $\widetilde{N}_G$ converges in distribution to the normal distribution by Theorem 1 in Janson (1988). We note that when $\alpha = 1$, (29) is consistent with (24).

Proposition 6.5 (Sparse regime). *Let G be a connected graph with $|V(G)| = r$ vertices, $r \geq 2$, satisfying Assumption 6.1 for all $n \geq 1$ in the sparse regime (20).*

1. *If G is a tree, i.e. $|E(G)| = r-1$, we have the cumulant bounds*

$$(K_1)^r \lambda^{\alpha-(\alpha-1)r} \leq \kappa_n(N_G) \leq n!^r (K_2)^r \lambda^{\alpha-(\alpha-1)r}, \quad \lambda \geq 1, \tag{25}$$

for some constants $K_1 > 0$, $K_2 > 1$ independent of $\lambda, n \geq 1$.

2. *If G is not a tree, i.e. $|E(G)| \geq r$, we have the cumulant bounds*

$$(K_1)^r \lambda^{r-\alpha|E(G)|} \leq \kappa_n(N_G) \leq n!^r (K_2)^r \lambda^{r-\alpha|E(G)|}, \quad \lambda \geq 1, \tag{26}$$

for some constants $K_1 > 0$, $K_2 > 1$ independent of $\lambda, n \geq 1$.

3. *If G is a cycle, i.e. $|E(G)| = r$, we have the cumulant bounds*

$$(K_1)^r \lambda^{-(\alpha-1)r} \leq \kappa_n(N_G) \leq n!^r (K_2)^r \lambda^{-(\alpha-1)r}, \quad \lambda \geq 1, \tag{27}$$

for some constants $K_1 > 0$, $K_2 > 1$ independent of $\lambda, n \geq 1$.

Proof. In the sparse regime (20), every edge in the graph ρ_G contributes a power $\lambda^{-\alpha}$ and every vertex contributes a power λ, hence every term in (16) contributes a power

$$\lambda^{|V(\rho_G)|-\alpha|E(\rho_G)|} = \lambda^{\alpha-(\alpha-1)|V(\rho_G)|+(|V(\rho_G)|-|E(\rho_G)|-1)\alpha} \leq \lambda^{\alpha-(\alpha-1)|V(\rho_G)|} \tag{28}$$

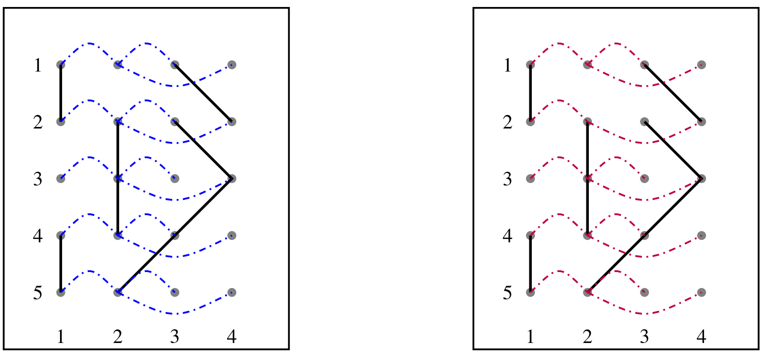
since $|V(\rho_G)| - |E(\rho_G)| - 1 \leq 0$. In addition, for any connected partition $\rho \in \Pi_{\widehat{1}}([n] \times [r])$, we have

$$r \leq |V(\rho_G)| \leq 1 + (r-1)n.$$

a) When G is a tree and the graph ρ_G is also a tree, i.e. $|V(\rho_G)| - |E(\rho_G)| - 1 = 0$, and the maximal order $\lambda^{\alpha-(\alpha-1)|V(\rho_G)|}$ is attained in (28), see Figure 7 for an example.
In this case, the corresponding term in (16) contributes a power

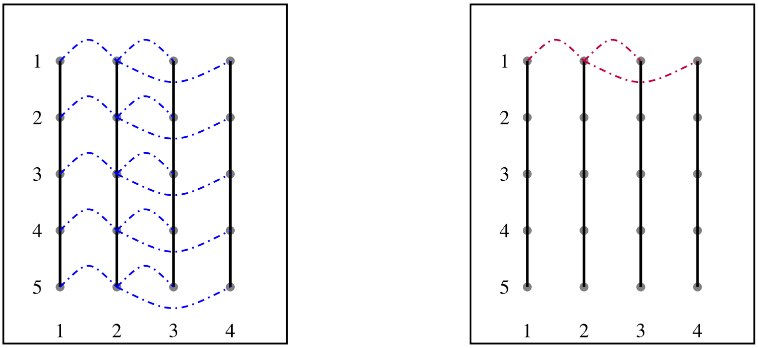
$$\lambda^{|V(\rho_G)|-\alpha|E(\rho_G)|} = \lambda^{\alpha-(\alpha-1)|V(\rho_G)|}, \qquad \lambda \geq 1.$$

In this case, since $|V(\rho_G)| \geq r$ and $\alpha \geq 1$, the optimal rate $\lambda^{\alpha-(\alpha-1)r}$ is attained by the partition diagrams $\Gamma(\rho,\pi)$ such that $|V(\rho_G)| = r$, as illustrated in Figure 8.



(a) Diagram $\Gamma(\rho,\pi)$ and multigraph $\widetilde{\rho}_G$ in blue. (b) Diagram $\Gamma(\sigma,\pi)$ and graph ρ_G in red.

Figure 7. Example of connected partition diagram with ρ_G a tree and $n = 5$, $r = 4$.



(a) Diagram $\Gamma(\rho,\pi)$ and multigraph $\widetilde{\rho}_G$ in blue. (b) Diagram $\Gamma(\rho,\pi)$ and graph ρ_G in red.

Figure 8. Tree diagram ρ_G with G a tree with $|V(\rho_G)| = r$ and $n = 5$, $r = 4$.

We conclude to (25) as in the proof of Proposition 6.3, by upper bounding the count of connected non-flat partitions from (5) and by lower bounding it by 1.

b) When G is not a tree it contains at least one cycle, and for any partition $\rho \in \Pi_{\widehat{1}}([n]\times[r])$ the same holds for the graph ρ_G. In this case, the highest order contribution in (16) is attained by connected non-flat partition diagrams $\Gamma(\rho,\pi)$, $\rho \in \Pi_{\widehat{1}}([n]\times[r])$, such that ρ_G has $|V(\rho_G)| = r$ vertices, and their contribution is given by a power of order $\lambda^{r-\alpha|E(G)|}$. An example of such partition ρ is given in Figure 9, with G a cycle.

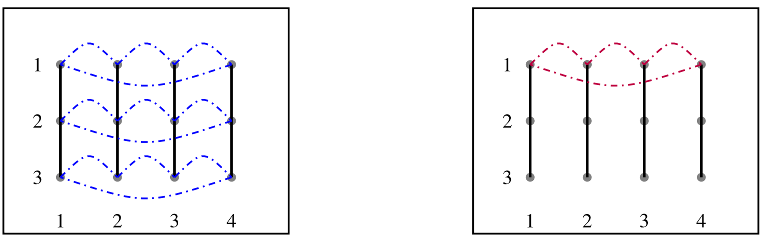
Indeed, in order to remain non-flat, the partition ρ can only be modified into a partition σ by splitting a block of ρ_G in two, which entails the addition of a number q of edges, $q \geq 1$, resulting into an additional factor $\lambda^{1-q\alpha} \leq 1$ that may only lower the order of the contribution, see Figures 10-13 for examples with G a graph containing one cycle.

When G is a triangle with $n = 2$ and $r = 3$, the above procedure can be reversed by first merging a vertex and then gluing edges, see Figure 14, which results into "overlapping" all copies of the graph G.

As in part (*b*) above, we lower bound $\kappa_n(N_G)$ using a single partition, and we upper bound using the total count of connected non-flat partitions using Lemma 2.8-*b*) to obtain (26).

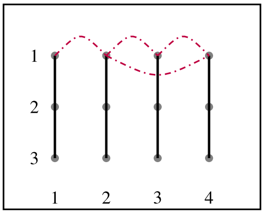
c) is a direct consequence of part (*b*) above. □

Corollary 6.6 (Sparse regime). *Let G be a connected graph with $|V(G)| = r$ vertices, $r \geq 2$, satisfying Assumption 6.1 for $n = 2$ in the sparse regime (20).*

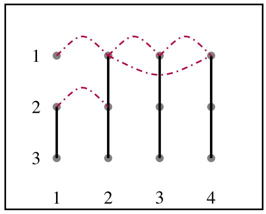


(a) Diagram $\Gamma(\rho,\pi)$ and multigraph $\widetilde{\rho}_G$ in blue. (b) Diagram $\Gamma(\sigma,\pi)$ and graph ρ_G in red.

Figure 9. Cycle graph ρ_G with G a cycle graph and $n = 5$, $r = 4$.

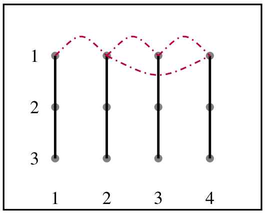


(a) Diagram $\Gamma(\rho, \pi)$ with order $\lambda^{4-4\alpha}$.

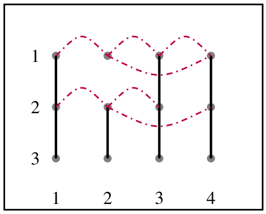


(b) Diagram $\Gamma(\sigma, \pi)$ with order $\lambda^{5-5\alpha} = \lambda^{4-4\alpha}\lambda^{-(\alpha-1)}$.

Figure 10. Splitting of a vertex with addition of one edge and $n = 3$, $r = 4$.

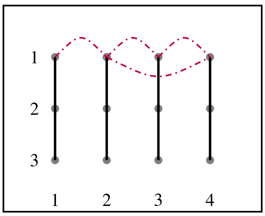


(a) Diagram $\Gamma(\rho, \pi)$ with order $\lambda^{4-4\alpha}$.

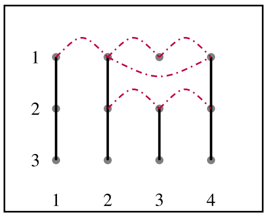


(b) Diagram $\Gamma(\sigma, \pi)$ with order $\lambda^{5-6\alpha} = \lambda^{4-4\alpha}\lambda^{1-2\alpha}$.

Figure 11. Splitting of a vertex with addition of three edges and $n = 3$, $r = 4$.

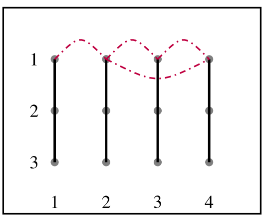


(a) Diagram $\Gamma(\rho, \pi)$ with order $\lambda^{4-4\alpha}$.

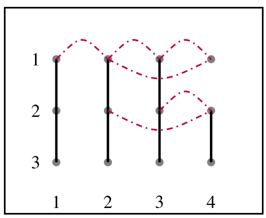


(b) Diagram $\Gamma(\sigma, \pi)$ with order $\lambda^{5-6\alpha} = \lambda^{4-4\alpha}\lambda^{1-2\alpha}$.

Figure 12. Splitting of a vertex with addition of two edges and $n = 3$, $r = 4$.

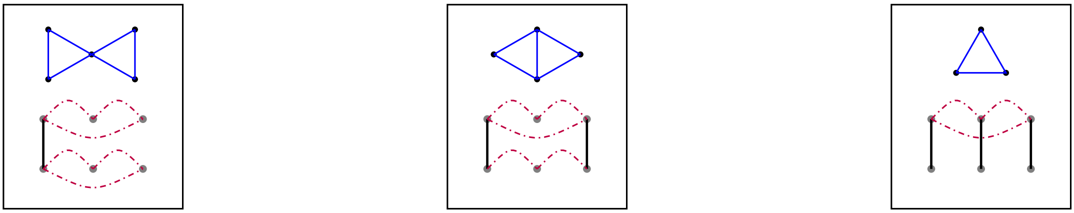


(a) Diagram $\Gamma(\rho, \pi)$ with order $\lambda^{4-4\alpha}$.



(b) Diagram $\Gamma(\sigma, \pi)$ with order $\lambda^{5-6\alpha} = \lambda^{4-4\alpha}\lambda^{1-2\alpha}$.

Figure 13. Splitting of a vertex with addition of two edges and $n = 3$, $r = 4$.



(a) Merging one vertex. (b) Gluing one edge. (c) Gluing three edges.

Figure 14. Diagram patterns with G a triangle and $n = 2$, $r = 3$.

1. *If G is a tree, i.e. $|E(G)| = r - 1$, we have the normalize cumulant bounds*

$$\left|\kappa_n(\widetilde{N}_G)\right| \le (K_3)^n n!^r \lambda^{-(\alpha-(\alpha-1)r)(n/2-1)}, \qquad \lambda \ge 1, \quad n \ge 2, \tag{29}$$

where $K_3 := \max((K_2)^r, 1)/(K_1)^{r/2}$.

2. *If G is not a tree, i.e. $|E(G)| \ge r$, we have the normalized cumulant bounds*

$$\left|\kappa_n(\widetilde{N}_G)\right| \le n!^r (K_3)^n \lambda^{(\alpha|E(G)|-r)(n/2-1)}, \qquad \lambda \ge 1, \quad n \ge 2, \tag{30}$$

for some $K_3 > 0$.

3. *If G is a cycle, i.e. $|E(G)| = r$, we have the normalized cumulant bounds*

$$\left|\kappa_n(\widetilde{N}_G)\right| \le n!^r (K_3)^n \lambda^{(\alpha-1)(n/2-1)r}, \qquad \lambda \ge 1, \quad n \ge 2, \tag{31}$$

for some $K_3 > 0$.

Proof. We note that the upper bound in (21) does not require Assumption 6.1. Regarding (29), we have

$$\begin{aligned}\left|\kappa_n(\widetilde{N}_G)\right| &\le \frac{n!^r (K_2)^r}{((K_1)^r \lambda^{\alpha-(\alpha-1)r})^{n/2}} \lambda^{\alpha-(\alpha-1)r} \\ &= \frac{(K_2)^r}{(K_1)^{nr/2}} n!^r \lambda^{-(\alpha-(\alpha-1)r)(n/2-1)}, \qquad n \ge 2.\end{aligned}$$

Regarding (30), we have

$$\left|\kappa_n(\widetilde{N}_G)\right| \le \frac{n!^r (K_2)^r \lambda^{r-\alpha|E(G)|}}{((K_1)^r \lambda^{r-\alpha|E(G)|})^{n/2}} = n!^r \frac{(K_2)^r}{(K_1)^{nr/2}} \lambda^{-(r-\alpha|E(G)|)(n/2-1)}, \quad n \ge 2.$$

Finally, the bound (31) is a direct consequence of (30). □

7. Asymptotic normality of subgraph counts

In this section, we let $H(x, y)$ be a connection function satisfying Assumption 6.1, and consider the random-connection model $G_{H_\lambda}(\Xi)$ where $H_\lambda(x, y) := c_\lambda H(x, y)$, $\lambda > 0$.

Corollary 7.1 (Dilute regime). *Let G be a connected graph with $|V(G)| = r$ vertices, $r \geq 2$, satisfying Assumption 6.1. In the dilute regime (18), the normalized subgraph count $\widetilde{N}_G$ in $G_{H_\lambda}(\Xi)$ satisfies the Kolmogorov distance bound*

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}\big(\widetilde{N}_G \leq x\big) - \Phi(x)\right| \leq C\lambda^{-1/(4r-2)}, \tag{32}$$

with rate exponent $1/(4r-2)$ as λ tends to infinity, where $C > 0$ depends only on H and G.

Proof. In the dilute regime, the cumulant bound (24) shows that the centered and normalized subgraph count $\widetilde{N}_G$ satisfies the Statulevičius condition (41) in the appendix, see Döring, Jansen and Schubert (2022), Rudzkis, Saulis and Statulevičius (1978), with $\gamma := r - 1$. We conclude by applying Corollary 6.4 and Lemma A.5-*i*) with $\gamma := r - 1$ and $\Delta_\lambda := \sqrt{K\lambda}$. □

In the sparse regime we have the following result, in which (33) is consistent with (32) when $\alpha = 1$.

Corollary 7.2 (Sparse regime). *Let G be a tree with $|V(G)| = r \geq 2$ vertices, satisfying Assumption 6.1. In the sparse regime (20) with $\alpha \in [1, r/(r-1))$, the normalized subgraph count $\widetilde{N}_G$ in $G_{H_\lambda}(\Xi)$ satisfies the Kolmogorov distance bound*

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}\big(\widetilde{N}_G \leq x\big) - \Phi(x)\right| \leq C\lambda^{-(\alpha-(\alpha-1)r)/(4r-2)}, \tag{33}$$

as λ tends to infinity, where $C > 0$ depends only on H and G.

Proof. This is a consequence of Corollary 6.6-*a*) and Lemma A.5-*i*) in the appendix, with $\gamma := r - 1$ and $\Delta_\lambda := (K\lambda)^{-(\alpha-(\alpha-1)r)/2}$ and $\alpha \in [1, r/(r-1))$. □

We note that up to division by $2r - 1$, the rate in (33) is consistent with the rate exponent $(\alpha - (\alpha-1)r)/2$ obtained for the counting of trees in the Erdős-Rényi graph, cf. Corollary 4.10 of Privault and Serafin (2020). In addition, since $(\alpha|E(G)| - r)(n/2 - 1) \geq (\alpha - 1)(n/2 - 1)r \geq 0$, no significant Kolmogorov bounds are derived from (30) and (31) for cycle and other non-tree graphs in the sparse regime, which is consistent with Corollaries 4.8-4.9 of Privault and Serafin (2020).

Taking $\Delta_\lambda = \sqrt{K\lambda}$, by Lemma A.5-*ii*) in the appendix, see Theorem 1.1 of Döring and Eichelsbacher (2013), we have the following result.

Corollary 7.3. *Let G be a connected graph with $|V(G)| = r \geq 2$ vertices, satisfying Assumption 6.1. The normalized subgraph count $\widetilde{N}_G$ satisfies a moderate deviation principle in the dilute regime of Corollary 7.1, with speed $a_\lambda^2 = o(\lambda^{1/(2r-1)})$ and rate function $x^2/2$.*

In addition, by Lemma A.5-*iii*) in the appendix, see the corollary of (Saulis and Statulevičius, 1991, Lemma 2.4), there exists a constant $K > 0$ such that for any sufficiently large λ we have the concentration inequality

$$\mathbb{P}\big(\big|\widetilde{N}_G\big| \geq x\big) \leq 2\exp\left(-\frac{1}{4}\min\left(\frac{x^2}{2^r}, \big(x\sqrt{K\lambda}\big)^{1/r}\right)\right), \quad x \geq 0, \tag{34}$$

in agreement with the rate in Theorem 1.1 of Bachmann and Reitzner (2018), which is stated for subgraph counts in random geometric graphs.

8. Subgraph counts in random geometric graphs

In this section, we consider subgraph counts in the (Poisson) random geometric graph model. Assume that μ is the Lebesgue measure, and that the intensity measure Λ takes the form

$$\Lambda(\mathrm{d}x) := \mathbf{1}_A(x)\mu(\mathrm{d}x), \qquad \lambda > 0,$$

where A is a Borel subset of $\mathbb{R}^d$ such that $\mu(A) < \infty$.

Definition 8.1. For every $\lambda > 0$, let $G_{H_\lambda}(\Xi)$ denote the random-connection model with connection function

$$H_\lambda(x,y) := \mathbf{1}_{\{\|x-y\| \le R_\lambda\}}, \qquad x,y \in \mathbb{R}^d,$$

for some function $R_\lambda > 0$ of λ. We consider the following regimes.

- Dense regime: we have

$$\lim_{\lambda\to\infty} \lambda R_\lambda^d \in (0,\infty].$$

- Sparse regime: we have

$$\lim_{\lambda\to\infty} \lambda R_\lambda^d = 0 \text{ and } \lim_{\lambda\to\infty} \lambda\big(\lambda R_\lambda^d\big)^{r-1} = \infty.$$

When $\lim_{\lambda\to\infty} \lambda R_\lambda^d = c \in (0,\infty)$, we also say that we are in the thermodynamic regime. The following result extends Proposition 3.2 of Lachièze-Rey and Peccati (2013) from second order cumulants to cumulants of any order.

Proposition 8.2. *Let G be a connected graph with $|V(G)| = r$ vertices, $r \ge 2$. In the random geometric graph model we have the following cumulant bounds.*

1. *(Dense regime). We have*

$$K_1(n-1)!\lambda^{1+(r-1)n}(R_\lambda^d)^{(r-1)n} \le \kappa_n(N_G) \le K_2 n!^r r!^{n-1}\lambda^{1+(r-1)n}(R_\lambda^d)^{(r-1)n}, \quad \lambda \ge 1, \tag{35}$$

   for some constants $K_1, K_2 > 0$ independent of $\lambda, n \ge 1$.
2. *(Thermodynamic regime). We have*

$$K_1\lambda \le \kappa_n(N_G) \le K_2 n!^r r!^{n-1}\lambda, \quad \lambda \ge 1, \tag{36}$$

   for some constants $K_1, K_2 > 0$ independent of $\lambda, n \ge 1$.
3. *(Sparse regime). We have*

$$K_1\lambda^r(R_\lambda^d)^{r-1} \le \kappa_n(N_G) \le K_2 n!^r r!^{n-1}\lambda^r(R_\lambda^d)^{r-1}, \quad \lambda \ge 1, \tag{37}$$

   for some constants $K_1, K_2 > 0$ independent of $\lambda, n \ge 1$.

Proof. By Proposition 5.1, letting $\bar\rho$ denote a spanning tree contained in ρ, we have

$$\kappa_n(N_G) = \sum_{\substack{\rho\in\Pi_{\widehat{1}}([n]\times[r]) \\ \rho\wedge\pi=\widehat{0}}} \lambda^{|\rho|}\int_{A^{|\rho|}}\left(\prod_{\{i,j\}\in E(\rho_G)} \mathbf{1}_{\{\|x_i-x_j\|\le R_\lambda\}}\right)\mathrm{d}x_1\cdots\mathrm{d}x_{|\rho|}$$

$$\le \sum_{\substack{\rho\in\Pi_{\widehat{1}}([n]\times[r])\\ \rho\wedge\pi=\widehat{0}}} \lambda^{|\rho|}\int_{A^{|\rho|}}\left(\prod_{\{i,j\}\in E(\bar\rho_G)}\mathbf{1}_{\{\|x_i-x_j\|\le R_\lambda\}}\right)\mathrm{d}x_1\cdots\mathrm{d}x_{|\rho|}$$
$$= \sum_{\substack{\rho\in\Pi_{\widehat{1}}([n]\times[r])\\ \rho\wedge\pi=\widehat{0}}} \lambda^{|\rho|}\mu(A)(v_dR_\lambda^d)^{|\rho|-1}.$$

a) In the dense regime with $\lim_{\lambda\to\infty}\lambda R_\lambda^d=\infty$, the dominating asymptotic order $\lambda(\lambda R_\lambda^d)^{(r-1)n}$ of $\kappa_n(N_G)$ is achieved when $|\rho|=1+(r-1)n$, which yields the upper bound in (35).
b) In the thermodynamic regime with $\lim_{\lambda\to\infty}\lambda R_\lambda^d=c>0$, the dominating asymptotic order of $\kappa_n(N_G)$ is λ, which yields the upper bound in (36).
c) In the sparse regime with $\lim_{\lambda\to\infty}\lambda R_\lambda^d=0$ and $\lim_{\lambda\to\infty}\lambda(\lambda R_\lambda^d)^{r-1}=\infty$, the dominating asymptotic order $\lambda(\lambda R_\lambda^d)^{r-1}$ of $\kappa_n(N_G)$ is achieved when $|\rho|=r$, which yields the upper bound in (37).
In addition, the kernel $H_\lambda(x,y)=\mathbf{1}_{\{\|x-y\|\le R_\lambda\}}$ satisfies Assumption 6.1 for all $n\ge1$, with $C_{H_\lambda}=v_d(R_\lambda/2)^d$ in the framework of above increasing intensity example, which similarly yields the lower bounds in (35)-(37). □

The next result is a direct consequence of Proposition 8.2.

Corollary 8.3. *Let G be a connected graph with $|V(G)|=r$ vertices, $r\ge2$. In the random geometric graph model we have the following normalized cumulant bounds.*

1. *(Dense and thermodynamic regime). We have*
$$\kappa_n\big(\widetilde N_G\big)\le n!^r(K\lambda)^{-(n/2-1)},\quad \lambda\ge1, \tag{38}$$
for some $K>0$ constant independent of $\lambda,n\ge1$.
2. *(Sparse regime). We have*
$$\kappa_n\big(\widetilde N_G\big)\le K\big(\lambda^rR_\lambda^{(r-1)d}\big)^{-(n/2-1)},\quad \lambda\ge1, \tag{39}$$
for some constants $K_1,K_2>0$ independent of $\lambda,n\ge1$.

The following result then follows from Corollary 8.3.

Corollary 8.4. *Let G be a connected graph with $|V(G)|=r$ vertices, $r\ge2$, in the random geometric graph model.*

1. *Dense/thermodynamic regimes. If* $\lim_{\lambda\to\infty}(\lambda R_\lambda^d)\in(0,\infty]$, *we have*
$$\sup_{x\in\mathbb{R}}\big|\mathbb{P}\big(\widetilde N_G\le x\big)-\Phi(x)\big|\le C\lambda^{-1/(4r-2)}.$$
2. *Sparse regime. If* $\lim_{\lambda\to\infty}\lambda R_\lambda^d=0$ *and* $\lim_{\lambda\to\infty}\lambda(\lambda R_\lambda^d)^{r-1}=\infty$, *we have*
$$\sup_{x\in\mathbb{R}}\big|\mathbb{P}\big(\widetilde N_G\le x\big)-\Phi(x)\big|\le C\big(\lambda^rR_\lambda^{(r-1)d}\big)^{-1/(4r-2)}. \tag{40}$$

Proof. In both cases (*i*) and (*ii*) we apply Corollary 8.3 and Lemma A.5-*i*) with $\gamma:=r-1$, by taking $\Delta_\lambda:=\sqrt{\lambda}$ in the dense and thermodynamic regimes, and $\Delta_\lambda:=\lambda^{r/2}R_\lambda^{(r-1)d/2}$ in the sparse regime. □

We note that Berry-Esseen convergence rates have been obtained for certain random functionals in the random geometric graph model, including total edge lengths in (Schulte, 2016, Corollary 4.3), clique counts using Poisson U-statistics in (Reitzner and Schulte, 2013, Theorem 4.1) and using stabilizing functionals in (Lachièze-Rey, Schulte and Yukich, 2019, Theorem 3.15), and k-hop counts in the one-dimensional unit disk model in (Privault, 2022, Proposition 8.1).

Moderate deviation and concentration inequalities

Letting $\gamma = r - 1$, In the dense and thermodynamic regimes of Corollary 8.4 with $\Delta_\lambda = \sqrt{\lambda}$, $\widetilde{N}_G / a_\lambda$ satisfies a moderate deviation principle with rate function $x^2/2$ and speed $a_\lambda^2 = o(\lambda^{1/(2r-1)})$ in the setting of Lemma A.5-*ii*) in the appendix, and the concentration inequality (34) holds by Lemma A.5-*iii*).

Appendix

The following results are summarized from the "main lemmas" in Chapter 2 of Saulis and Statulevičius (1991) and Döring and Eichelsbacher (2013), and are tailored for our applications to the random-connection model.

Lemma A.5. *Let $(X_\lambda)_{\lambda \geq 1}$ be a family of random variables with mean zero and unit variance for all $\lambda > 0$. Suppose that for all $\lambda \geq 1$, all moments of the random variable X_λ exist and that the cumulants of X_λ satisfy*

$$|\kappa_n(X_\lambda)| \leq \frac{(n!)^{1+\gamma}}{(\Delta_\lambda)^{n-2}}, \qquad n \geq 3, \tag{41}$$

where $\gamma \geq 0$ is a constant not depending on λ, while $\Delta_\lambda \in (0, \infty)$ may depend on λ. Then, the following assertions hold.

1. *(Kolmogorov bound, (Saulis and Statulevičius, 1991, Corollary 2.1) and (Döring, Jansen and Schubert, 2022, Theorem 2.4)) One has*

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(X_\lambda \leq x) - \Phi(x)| \leq \frac{C}{(\Delta_\lambda)^{1/(1+2\gamma)}}, \tag{42}$$

   for some constant $C > 0$ depending only on γ.
2. *(Moderate deviation principle, (Döring and Eichelsbacher, 2013, Theorem 1.1) and (Döring, Jansen and Schubert, 2022, Theorem 3.1)). Let $(a_\lambda)_{\lambda > 0}$ be a sequence of real numbers tending to infinity, and such that*

$$\lim_{\lambda \to \infty} \frac{a_\lambda}{(\Delta_\lambda)^{1/(1+2\gamma)}} = 0.$$

   Then, $(a_\lambda^{-1} X_\lambda)_{\lambda > 0}$ satisfies a moderate deviation principle with speed a_λ^2 and rate function $x^2/2$.
3. *(Concentration inequality, corollary of (Saulis and Statulevičius, 1991, Lemma 2.4) and (Döring, Jansen and Schubert, 2022, Theorem 2.5)). For any sufficiently large λ,*

$$\mathbb{P}(|X_\lambda| \geq x) \leq 2 \exp\left(-\frac{1}{4} \min\left(\frac{x^2}{2^{1+\gamma}}, (x\Delta_\lambda)^{1/(1+\gamma)} \right) \right), \quad x \geq 0. \tag{43}$$

Acknowledgments

We thank M. Schulte and C. Thäle for a correction to an earlier version of Lemma 2.8, and the anonymous referees for useful suggestions. This research is supported by the Ministry of Education, Singapore, under its Tier 2 Grant MOE-T2EP20120-0005.

References

Bachmann, S. and Reitzner, M. (2018). Concentration for Poisson U-statistics: Subgraph counts in random geometric graphs. *Stochastic Process. Appl.* **128** 3327–3352. MR3849811 https://doi.org/10.1016/j.spa.2017.11.001

Balakrishnan, R. and Ranganathan, K. (2012). *A Textbook of Graph Theory*, 2nd ed. *Universitext*. New York: Springer. MR2977757 https://doi.org/10.1007/978-1-4614-4529-6

Barbour, A.D., Karoński, M. and Ruciński, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* **47** 125–145. MR1047781 https://doi.org/10.1016/0095-8956(89)90014-2

Bender, E.A., Odlyzko, A.M. and Richmond, L.B. (1985). The asymptotic number of irreducible partitions. *European J. Combin.* **6** 1–6. MR0793481 https://doi.org/10.1016/S0195-6698(85)80015-9

Bogdan, K., Rosiński, J., Serafin, G. and Wojciechowski, Ł. (2017). Lévy systems and moment formulas for mixed Poisson integrals. In *Stochastic Analysis and Related Topics*. *Progress in Probability* **72** 139–164. Cham: Birkhäuser/Springer. MR3737628 https://doi.org/10.1007/978-3-319-59671-6_7

Can, V.H. and Trinh, K.D. (2022). Random connection models in the thermodynamic regime: Central limit theorems for add-one cost stabilizing functionals. *Electron. J. Probab.* **27** 1–40. MR4388562 https://doi.org/10.1214/22-EJP759

Döring, H. and Eichelsbacher, P. (2013). Moderate deviations via cumulants. *J. Theoret. Probab.* **26** 360–385. MR3055808 https://doi.org/10.1007/s10959-012-0437-0

Döring, H., Jansen, S. and Schubert, K. (2022). The method of cumulants for the normal approximation. *Probab. Surv.* **19** 185–270. MR4408127 https://doi.org/10.1214/22-ps7

Eichelsbacher, P. and Rednoss, B. (2023). Kolmogorov bounds for decomposable random variables and subgraph counting by the Stein-Tikhomirov method. *Bernoulli* **29** 1821–1848. MR4580898 https://doi.org/10.3150/22-bej1522

Eichelsbacher, P. and Thäle, C. (2014). New Berry-Esseen bounds for non-linear functionals of Poisson random measures. *Electron. J. Probab.* **19** 1–25. MR3275854 https://doi.org/10.1214/EJP.v19-3061

Erdős, P. and Rényi, A. (1959). On random graphs. I. *Publ. Math. Debrecen* **6** 290–297. MR0120167 https://doi.org/10.5486/pmd.1959.6.3-4.12

Féray, V., Méliot, P.-L. and Nikeghbali, A. (2016). *Mod-ϕ Convergence: Normality Zones and Precise Deviations*. *SpringerBriefs in Probability and Mathematical Statistics*. Cham: Springer. MR3585777 https://doi.org/10.1007/978-3-319-46822-8

Gilbert, E.N. (1959). Random graphs. *Ann. Math. Stat.* **30** 1141–1144. MR0108839 https://doi.org/10.1214/aoms/1177706098

Götze, F., Heinrich, L. and Hipp, C. (1995). m-dependent random fields with analytic cumulant generating function. *Scand. J. Stat.* **22** 183–195. MR1339750

Grote, J. and Thäle, C. (2018a). Concentration and moderate deviations for Poisson polytopes and polyhedra. *Bernoulli* **24** 2811–2841. MR3779703 https://doi.org/10.3150/17-BEJ946

Grote, J. and Thäle, C. (2018b). Gaussian polytopes: A cumulant-based approach. *J. Complexity* **47** 1–41. MR3804598 https://doi.org/10.1016/j.jco.2018.03.001

Gusakova, A. and Thäle, C. (2021). The volume of simplices in high-dimensional Poisson-Delaunay tessellations. *Ann. Henri Lebesgue* **4** 121–153. MR4213157 https://doi.org/10.5802/ahl.68

Heinrich, L. (2007). An almost-Markov-type mixing condition and large deviations for Boolean models in the line. *Acta Appl. Math.* **96** 247–262. MR2327539 https://doi.org/10.1007/s10440-007-9105-2

Heinrich, L. and Spiess, M. (2009). Berry-Esseen bounds and Cramér-type large deviations for the volume distribution of Poisson cylinder processes. *Lith. Math. J.* **49** 381–398. MR2591874 https://doi.org/10.1007/s10986-009-9061-9

Jansen, S. (2019). Cluster expansions for Gibbs point processes. *Adv. in Appl. Probab.* **51** 1129–1178. MR4032174 https://doi.org/10.1017/apr.2019.46

Janson, S. (1988). Normal convergence by higher semi-invariants with applications to sums of dependent random variables and random graphs. *Ann. Probab.* **16** 305–312. MR0920273

Khorunzhiy, O. (2008). On connected diagrams and cumulants of Erdős-Rényi matrix models. *Comm. Math. Phys.* **282** 209–238. MR2415478 https://doi.org/10.1007/s00220-008-0533-2

Krokowski, K., Reichenbachs, A. and Thäle, C. (2017). Discrete Malliavin-Stein method: Berry-Esseen bounds for random graphs and percolation. *Ann. Probab.* **45** 1071–1109. MR3630293 https://doi.org/10.1214/15-AOP1081

Lachièze-Rey, R. and Peccati, G. (2013). Fine Gaussian fluctuations on the Poisson space II: Rescaled kernels, marked processes and geometric U-statistics. *Stochastic Process. Appl.* **123** 4186–4218. MR3096352 https://doi.org/10.1016/j.spa.2013.06.004

Lachièze-Rey, R. and Reitzner, M. (2016). U-statistics in stochastic geometry. In *Stochastic Analysis for Poisson Point Processes*. *Bocconi Springer Ser.* **7** 229–253. Bocconi Univ. Press. MR3585402

Lachièze-Rey, R., Schulte, M. and Yukich, J.E. (2019). Normal approximation for stabilizing functionals. *Ann. Appl. Probab.* **29** 931–993. MR3910021 https://doi.org/10.1214/18-AAP1405

Last, G., Nestmann, F. and Schulte, M. (2021). The random connection model and functions of edge-marked Poisson processes: Second order properties and normal approximation. *Ann. Appl. Probab.* **31** 128–168. MR4254476 https://doi.org/10.1214/20-aap1585

Last, G., Peccati, G. and Schulte, M. (2016). Normal approximation on Poisson spaces: Mehler's formula, second order Poincaré inequalities and stabilization. *Probab. Theory Related Fields* **165** 667–723. MR3520016 https://doi.org/10.1007/s00440-015-0643-7

Last, G. and Penrose, M. (2018). *Lectures on the Poisson Process*. *Institute of Mathematical Statistics Textbooks* **7**. Cambridge: Cambridge Univ. Press. MR3791470

Malyshev, V.A. and Minlos, R.A. (1991). *Gibbs Random Fields: Cluster Expansions*. *Mathematics and Its Applications (Soviet Series)* **44**. Dordrecht: Kluwer Academic. Translated from the Russian by R. Kotecký and P. Holický. MR1191166 https://doi.org/10.1007/978-94-011-3708-9

Peccati, G. and Taqqu, M.S. (2011). *Wiener Chaos: Moments, Cumulants and Diagrams: A Survey with Computer Implementation*. *Bocconi & Springer Series* **1**. Milan: Springer; Bocconi Univ. Press. Supplementary material available online. MR2791919 https://doi.org/10.1007/978-88-470-1679-8

Penrose, M. (2003). *Random Geometric Graphs*. *Oxford Studies in Probability* **5**. Oxford: Oxford Univ. Press. MR1986198 https://doi.org/10.1093/acprof:oso/9780198506263.001.0001

Penrose, M.D. and Yukich, J.E. (2001). Central limit theorems for some graphs in computational geometry. *Ann. Appl. Probab.* **11** 1005–1041. MR1878288 https://doi.org/10.1214/aoap/1015345393

Penrose, M.D. and Yukich, J.E. (2005). Normal approximation in geometric probability. In *Stein's Method and Applications*. *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **5** 37–58. Singapore: Singapore Univ. Press. MR2201885 https://doi.org/10.1142/9789812567673_0003

Privault, N. (2012). Moments of Poisson stochastic integrals with random integrands. *Probab. Math. Statist.* **32** 227–239. MR3021456

Privault, N. (2019). Moments of k-hop counts in the random-connection model. *J. Appl. Probab.* **56** 1106–1121. MR4041451 https://doi.org/10.1017/jpr.2019.63

Privault, N. (2022). Asymptotic analysis of k-hop connectivity in the 1D unit disk random graph model. Preprint, 40 pages. Available at arXiv:2203.14535.

Privault, N. and Serafin, G. (2020). Normal approximation for sums of weighted U-statistics—application to Kolmogorov bounds in random subgraph counting. *Bernoulli* **26** 587–615. MR4036045 https://doi.org/10.3150/19-BEJ1141

Privault, N. and Serafin, G. (2022). Berry-Esseen bounds for functionals of independent random variables. *Electron. J. Probab.* **27** 1–37. MR4440064 https://doi.org/10.1214/22-ejp795

Reitzner, M. and Schulte, M. (2013). Central limit theorems for U-statistics of Poisson point processes. *Ann. Probab.* **41** 3879–3909. MR3161465 https://doi.org/10.1214/12-AOP817

Röllin, A. (2022). Kolmogorov bounds for the normal approximation of the number of triangles in the Erdős-Rényi random graph. *Probab. Engrg. Inform. Sci.* **36** 747–773. MR4459700 https://doi.org/10.1017/S0269964821000061

Rota, G.-C. (1964). On the foundations of combinatorial theory. I. Theory of Möbius functions. *Z. Wahrsch. Verw. Gebiete* **2** 340–368. MR0174487 https://doi.org/10.1007/BF00531932

Ruciński, A. (1988). When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields* **78** 1–10. MR0940863 https://doi.org/10.1007/BF00718031

Rudzkis, R., Saulis, L. and Statulevičius, V. A. (1978). A general lemma on probabilities of large deviations. *Liet. Mat. Rink.* **18** 99–116, 217. MR0501287

Saulis, L. and Statulevičius, V.A. (1991). *Limit Theorems for Large Deviations*. *Mathematics and Its Applications (Soviet Series)* **73**. Dordrecht: Kluwer Academic. Translated and revised from the 1989 Russian original. MR1171883 https://doi.org/10.1007/978-94-011-3530-6

Schulte, M. (2016). Normal approximation of Poisson functionals in Kolmogorov distance. *J. Theoret. Probab.* **29** 96–117. MR3463079 https://doi.org/10.1007/s10959-014-0576-6

Schulte, M. and Thäle, C. (2023). Moderate deviations on Poisson chaos. Preprint. Available at arXiv:2304:00876v1.

Zhang, Z.-S. (2022). Berry-Esseen bounds for generalized U-statistics. *Electron. J. Probab.* **27** 1–36. MR4492985 https://doi.org/10.1214/22-ejp860

Received June 2023 and revised December 2023