Gaussian fluctuations of generalized U-statistics and subgraph counting in the binomial random-connection model

Qingwei Liu* Nicolas Privault†

November 10, 2025

Abstract

We derive normal approximation bounds for generalized U-statistics of the form

$$S_{n,k}(f) := \sum_{\substack{1\leq \beta(1),\ldots,\beta(k)\leq n \\ \beta(i)\neq\beta(j),\ 1\leq i\neq j\leq k}} f\big(X_{\beta(1)},\ldots,X_{\beta(k)},Y_{\beta(1),\beta(2)},\ldots,Y_{\beta(k-1),\beta(k)}\big),$$

where $\{X_i\}_{1\leq i\leq n}$ and $\{Y_{i,j}\}_{1\leq i<j\leq n}$ are independent sequences of i.i.d. random variables. Our approach relies on moment identities and cumulant bounds that are derived using partition diagram arguments. Normal approximation bounds in the Kolmogorov distance and moderate deviation results are then obtained by the cumulant method. Those results are applied to subgraph counting in the binomial random-connection model, which is a generalization of the Erdős-Rényi model.

Keywords: Generalized U-statistics, binomial random-connection model, inhomogeneous random graph, subgraph count, cumulant method, dependency graph.

Mathematics Subject Classification: 60F05, 60G50, 05C80.

1 Introduction

Second-order U-statistics can be viewed as quadratic random functionals of the form

$$\sum_{1\leq i\neq j\leq n} Y_{i,j}X_iX_j \tag{1.1}$$

which are used to model potentials and partition functions in the framework of the Gaussian Unitary Ensemble in statistical mechanics, where $(Y_{i,j})_{1\leq i\neq j\leq n}$ is a possibly random adjacency

*School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia. Email: qingwei.liu@unimelb.edu.au

†Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371, Email: nprivault@ntu.edu.sg

matrix made of independent entries, and $(X_i)_{1\le i\le n}$ is a sequence of independent identically distributed random variables. Cumulant bounds of (1.1) have been obtained in [Kho08] when $(Y_{i,j})_{1\le i\ne j\le n}$ represents the adjacency matrix of the Erdős-Rényi random graph, and general approximation results in distribution have been recently obtained in [BDMM24] when $(Y_{i,j})_{1\le i\ne j\le n}$ is deterministic. On the other hand, more general pair interactions can be modeled using U-statistics of the form

$$\sum_{1\le i\ne j\le n} f(X_i, X_j),$$

while higher-order U-statistics can model nonlinear interactions in non-Gaussian frameworks.

In this paper, we derive normal approximation results and cumulant bounds for generalized U-statistics of the form

$$S_{n,k}(f) := \sum_{\substack{1\le \beta(1),\ldots,\beta(k)\le n \\ \beta(i)\ne\beta(j),\ 1\le i\ne j\le k}} f\big(X_{\beta(1)},\ldots,X_{\beta(k)},Y_{\beta(1),\beta(2)},\ldots,Y_{\beta(1),\beta(k)},Y_{\beta(2),\beta(3)},\ldots,Y_{\beta(k-1),\beta(k)}\big), \tag{1.2}$$

which have been introduced in [JN91] as a powerful tool for studying the normal and non-normal asymptotic distributions of subgraph counts in inhomogeneous random graphs. Here, $\{X_i\}_{1\le i\le n}$ and $\{Y_{i,j}\}_{1\le i<j\le n}$ are two independent sequences of i.i.d. random elements taking values respectively in a Borel space $\mathcal{S}$ and a measurable space $\mathcal{M}$, and $f : \mathcal{S}^k\times\mathcal{M}^{k(k-1)/2}\to\mathbb{R}$ is a measurable function, $k\ge 2$, with $Y_{i,j} = Y_{j,i}$ if $1\le j<i\le n$.

In Corollary 4.3, we obtain a Kolmogorov distance bound of the form

$$\sup_{x\in\mathbb{R}}\big|\mathbb{P}\left(\overline{S}_{n,k}(f)\le x\right)-\Phi(x)\big| \le \frac{C(f,k)}{n^{1/(2+4k)}}, \quad n\ge 4(k-1),$$

for the normalized generalized U-statistics

$$\overline{S}_{n,k}(f) := \frac{S_{n,k}(f)-\kappa_1(S_{n,k}(f))}{\sqrt{\kappa_2(S_{n,k}(f))}},$$

where Φ is the cumulative distribution function of the standard normal distribution and $C(f,k)>0$ depends only on f and on $k\ge 2$. Our approach relies on moment identities and cumulant bounds for generalized U-statistics established in Theorems 3.2 and 4.1. Berry-Esseen bounds for the normal approximation of general functionals of binomial point processes have been obtained in [LRP17], with application to U-statistics and set approximation for random tessellations. However, generalized U-statistics of the form (1.2) include an additional layer of randomness due to the random sequence $\{Y_{i,j}\}_{1\le i<j\le n}$. In Corollary 4.4 we obtain a moderate deviation result for the normalized U-statistics $\overline{S}_{n,k}(f)$.

Starting with Section 5, we apply our normal approximation results for generalized U-statistics to subgraph counting in the binomial random-connection model. Random-connection models (RCMs) are random graphs which are based on randomly located vertices which are independently connected with a location-dependent probability. As a generalization of the Erdős-Rényi model, the binomial RCM has gained significant attention and has been studied under different names, for example as inhomogeneous random graphs, c.f. [DF14, Pen18, HPČ21], and as graphon-based random graphs c.f. [CGR16, Zha22, BCJ23].

Distributional approximations for count statistics on random-connection models whose vertices are generated according to a Poisson point process, have been investigated in a number of recent works, including vertices counts [Pen18], component counts [LNS21], and subgraph counts [CT22, LP24a]. Recently, Poisson approximation with bounds for subgraph counts in general random-connection models have been derived in [LX25], and the cumulant method has been applied to subgraph counting weighted random connection models in [HHO25].

More formally, let $\mathcal{X}_n = \{X_1,\dots,X_n\}$ denote a family of i.i.d. random points with a common distribution μ on $\mathcal{S} = \mathbb{R}^d$, for some $n \ge 2$. Given $H : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ a symmetric measurable connection function, the binomial random-connection model with connection function H is the random graph on the binomial point process $\mathcal{X}_n = \{X_1,\dots,X_n\}$ constructed by adding edges independently with probability $H(X_i,X_j)$, to each distinct pair (X_i,X_j) of vertices, $1\le i \ne j \le n$.

When the connection function is taken as $H(x,y) = \mathbf{1}_{\{\|x-y\|\le r\}}$, $x,y\in\mathbb{R}^d$ for some $r>0$, vertices are connected in a deterministic way, and the binomial RCM becomes a random geometric graph, or Gilbert graph, c.f. [Pen03]. When $H(x,y)\equiv p_n$, $x,y\in\mathbb{R}^d$, the binomial RCM recovers the classical Erdős-Rényi random graph. In this case, subgraph counting in the binomial RCM is a natural extension of the subgraph counting problem in the Erdős-Rényi model, see [JŁR00, FMN16].

Although the asymptotic behavior of subgraph counts in the binomial RCM was studied in detail in [JN91, Jan97, BCJ23], convergence rates for the distributional approximation of subgraph counts have only been recently discussed, see [KR21, Zha22]. In [DST16], the Poisson approximation of standard (not generalized) U-statistics has been considered in the binomial model. More recently, general approximation results for standard second-order U-statistics have been obtained in [BDMM24]. However, none of those works, including [Zha22], consider the case where the probabilities of connecting two vertices tends to zero as n tends to infinity, as is typical in the Erdős-Rényi model.

In contrast with the centered subgraph counts considered in [KR21], we allow the connection probability of any distinct pair (X_i, X_j) of vertices in the binomial RCM to be of the form $p_n H(X_i, X_j)$ where $p_n \in (0,1)$, $1 \leq i \neq j \leq n$. We consider in particular the case where p_n may tend to zero as n tends to infinity, and study the corresponding normal approximation of subgraph counts.

In Theorem 5.2 we derive upper bounds on the cumulants of subgraph counts in the binomial random-connection model. Note that related cumulant bounds have been obtained in the Erdős-Rényi model in [Kho08] for the counts of line and cycle graphs, and in [FMN16] for general subgraph counts. In comparison with the Poisson random-connection model considered in [LP24a, LP24b], cumulants admit no simplified expression using sums over connected non-flat partitions in the binomial random-connection model. For this reason, cumulant bounds have to be derived using specific arguments.

Then, by combining Theorem 5.2 with variance lower bounds obtained in Proposition 6.1, in Theorem 7.5 we obtain cumulant growth rates for the counts of strongly balanced connected graphs. Our proof relies on dependency graph methods and the convex analysis of planar diagrams, which were introduced in [ŁR92] to study the behaviour of the variance of subgraph counts in the Erdős-Rényi model.

Cumulant growth rates for the normalized counts of strongly balanced connected graphs are then obtained in Corollary 7.8 under Assumption 4.1 in the case where $p_n = o(1)$, using the variance lower bound for subgraph counts established in Proposition 7.7. Note however that Assumption 4.1 is valid in the binomial RCM, and is not satisfied in the Erdős-Rényi model.

Using the cumulant method and the Statulevičius condition, see Appendix A, Kolmogorov distance rates of the form

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(\overline{N}_G \leq x) - \Phi(x)| \leq \begin{cases} \dfrac{C}{n^{1/(2+4v(G))}} & \textit{when} \quad n^{-(v(G)-1)/e(G)} \ll p_n, \\ \dfrac{C}{\left(n^{v(G)} p_n^{e(G)}\right)^{1/(2+4v(G))}} & \textit{when} \quad p_n \ll n^{-(v(G)-1)/e(G)}, \end{cases}$$

$n \geq 4(v(G)-1)$, are obtained in Corollary 7.9 for the normalized subgraph count

$$\overline{N}_G := \frac{N_G - \kappa_1(N_G)}{\sqrt{\kappa_2(N_G)}}$$

of a strongly balanced connected graph G with $v(G)$ vertices. This extends the results of [PS18] and [ER23] for the approximation of subgraph counts from the Erdős-Rényi model to the binomial RCM. Although the convergence rates in the Kolmogorov distance obtained in do not match the optimal rate obtained in [Zha22], they allow us to consider the case where $p_n = o(1)$. In Corollary 7.11, we investigate the threshold phenomenon for the containment of subgraphs in the binomial RCM.

In Corollary 7.10 we obtain moderate deviation results for the normalized subgraph count $\overline{N}_G$, see also [DE09, Theorem 1.1], [DE13, Theorem 2.3], [FMN16, Theorem 10.0.2], [AcdOG25, Theorem 1.5] for moderate deviation results in the Erdős-Rényi model. In Corollary 7.11, we investigate the threshold phenomenon for the containment of subgraphs in the binomial RCM.

The paper is organized as follows. In Section 2 we recall some notation and definitions related to set partitions and diagrams. In Section 3 we derive moment identities for generalized U-statistics. Section 4 gives cumulant bounds for generalized U-statistics and further obtains normal approximation results via the cumulant method. In Section 5, we obtain cumulant bounds for subgraph counts in the binomial random-connection model, which allows the connection probability between pairs of vertices to be significantly small. In Section 6, we derive lower bounds on the variance of subgraph counts in the binomial RCM. This is crucial for proving the results in Section 7. In Section 7 we obtain normal approximation for subgraph counts in the binomial RCM through a refined analysis on the cumulant growth rates and a threshold phenomenon for subgraph containment. In the Appendix A, we provide a brief review of the Statulevičius condition and its application to the cumulant method.

2 Set partitions and diagrams

In what follows, we let $[n] := \{1, 2, \ldots, n\}$ for $n \geq 1$, and let $\Pi(b)$ denote the collection of set partitions of any finite set b. Given two set partitions $\rho_1, \rho_2 \in \Pi(b)$, we say that ρ_1 is coarser than ρ_2 (i.e. ρ_2 is finer than ρ_1), and we write $\rho_2 \preceq \rho_1$, if and only if each block of ρ_2 is contained in a block of ρ_1. We use $\rho_1 \vee \rho_2$ for the finest partition which is coarser than both ρ_1 and ρ_2, and denote by $\rho_1 \wedge \rho_2$ the coarsest partition which is finer than both of ρ_1

and ρ_2. We also let $\widehat{1} := \{b\}$ denote the single-block coarsest partition of b, whereas $\widehat{0}$ stands for the partition made of singletons. Given $k \geq 2$ and $j \geq 1$, we let $\pi := \{\pi_1, \ldots, \pi_j\}$ denote the partition of $[j] \times [k]$ defined as

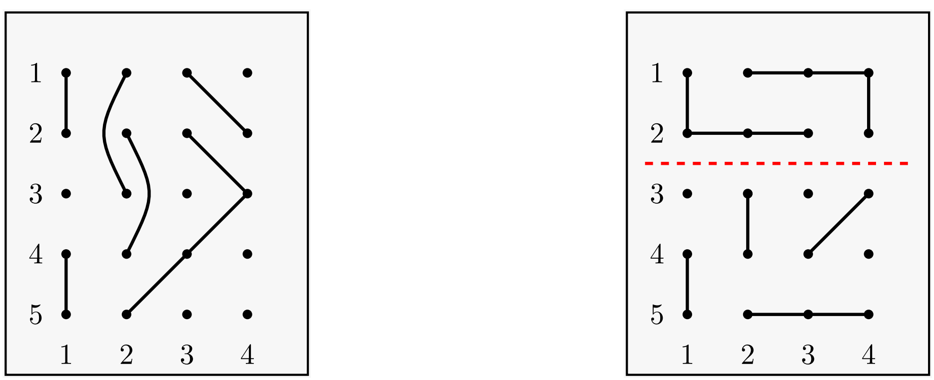
$$\pi_i := \{(i, \ell) : 1 \leq \ell \leq k\}, \quad i = 1, \ldots, j.$$

For $j, k \geq 1$ we also let $\pi_\eta := (\pi_i)_{i \in \eta} \in \Pi(\eta \times [k])$ denote the partition made of $|\eta|$ blocks of size k. A partition $\rho \in \Pi([j] \times [k])$ is said to be *non-flat* if $\rho \wedge \pi = \widehat{0}$, see Chapter 4 of [PT11] and Figure 1.

(a) Non-flat partition. (b) Flat partition.

Figure 1: Examples of partition diagrams with $n = 3$ and $r = 4$.

A partition $\rho \in \Pi([j] \times [k])$ is said to be *connected* if $\rho \vee \pi = \widehat{1}$, see Figure 2.



(a) Connected partition. (b) Non-connected partition.

Figure 2: Examples of partition diagrams with $n = 5$ and $r = 4$.

We also let $\Pi_{\widehat{1}}([j] \times [k])$ denote the collection of all connected partitions of $[j] \times [k]$, and denote by

$$\mathrm{CNF}(j, k) := \left\{\rho : \rho \in \Pi_{\widehat{1}}([j] \times [k]),\ \rho \wedge \pi = \widehat{0}\right\}$$

the set of all connected and non-flat partitions of $[j] \times [k]$, for $j, k \geq 1$. Let $[n]^k_{\neq}$ denote the collection of distinct k-fold indexes

$$[n]^k_{\neq} := \{\beta = (\beta(1), \ldots, \beta(k)) \in [n]^k : \beta(i) \neq \beta(j) \text{ for } 1 \leq i \neq j \leq k\}, \qquad k \geq 1.$$

The role of the partition $\sqcap(\alpha)$ introduced in Definition 2.1 is to group each set of identical entries in a family of k-tuples into a partition block. Later on, it will be used to identify the common random variables appearing in repeated copies of $(X_1, \ldots, X_n)$ for the computation of joint cumulants.

Definition 2.1 *Given a sequence*

$$\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_j \end{bmatrix} = \begin{bmatrix} \alpha_1(1) & \cdots & \alpha_1(k) \\ \vdots & \ddots & \vdots \\ \alpha_j(1) & \cdots & \alpha_j(k) \end{bmatrix} \in [n]^k_{\neq} \times \cdots \times [n]^k_{\neq},$$

we let $\sqcap(\alpha)$ denote the partition of $[j] \times [k]$ such that each block of $\sqcap(\alpha)$ is made of elements (i, ℓ) that correspond to a same value of $\alpha_i(\ell)$.

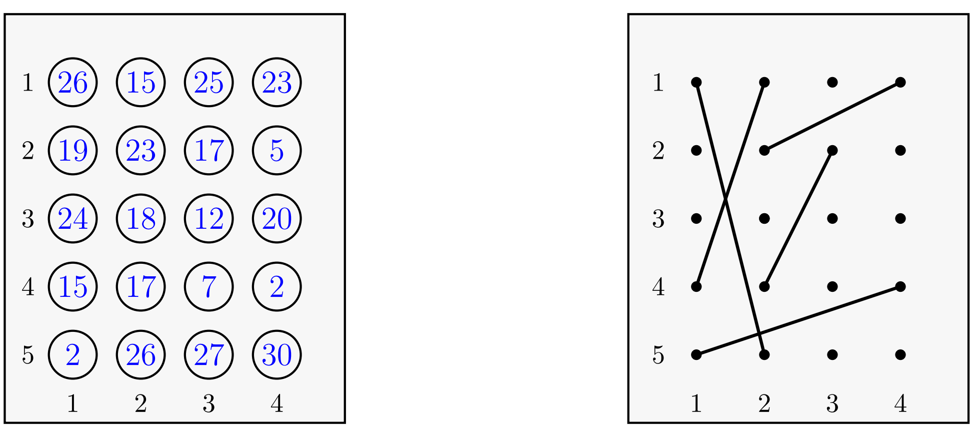
Next is an example of a sequence $\alpha \in [n]^k_{\neq} \times \cdots \times [n]^k_{\neq}$ and of the partition $\sqcap(\alpha)$ of $[j] \times [k]$ it generates.

Example 2.2 *Taking $n = 30$, $j = 5$, $k = 4$, and*

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix} = \begin{bmatrix} 26 & 15 & 25 & 23 \\ 19 & 23 & 17 & 5 \\ 24 & 18 & 12 & 20 \\ 15 & 17 & 7 & 2 \\ 2 & 26 & 27 & 30 \end{bmatrix},$$

the partition $\sqcap(\alpha)$ of $[5] \times [4]$ is given in Figure 3 by

$$\sqcap(\alpha) = \big\{\{(1,1),(5,2)\},\{(1,2),(4,1)\},\{(1,3)\},\{(1,4),(2,2)\},\{(2,1)\},\{(2,3),(4,2)\},$$
$$\{(2,4)\},\{(3,1)\},\{(3,2)\},\{(3,3)\},\{(3,4)\},\{(4,3)\},\{(4,4),(5,1)\},\{(5,3)\},\{(5,4)\}\big\}.$$



(a) $\alpha = [\alpha_1, \ldots, \alpha_5]^\top$. (b) Partition $\sqcap(\alpha)$.

Figure 3: Example for the mapping $\sqcap$ with $j = 5$ and $k = 4$.

We let $v(G) := |V_G|$ and $e(G) := |E_G|$ be the number of vertices and the number of edges of any graph $G = (V_G, E_G)$ with vertex set V_G and edge set E_G. A subgraph of G is a graph $H = (V_H, E_H)$ such that $V_H \subset V_G$ and $E_H \subset E_G$, and H is an induced subgraph of G, if E_H consists of all edges of G having both endpoints in V_H. Two graphs $G = (V_G, E_G)$ and $H = (V_H, E_H)$ are isomorphic if there is a bijection $T : V_G \to V_H$ such that $\{u, v\} \in E_G$ if and only if $\{T(u), T(v)\} \in E_H$ for any $u \neq v \in V_G$, in which case we write $H = T(G)$. The permutations α of V_G such that $\alpha(G) = G$ form a group called the automorphism group, and we let $a(G)$ denote the cardinality of this group.

Definition 2.3 *Consider $G_1, \ldots, G_j$ copies of a connected graph G with $v(G) = k \geq 2$ vertices, respectively built on $\pi_1, \ldots, \pi_j$, $j \geq 1$, and let $\rho \in \Pi([j] \times [k])$ be a partition of $[j] \times [k]$.*

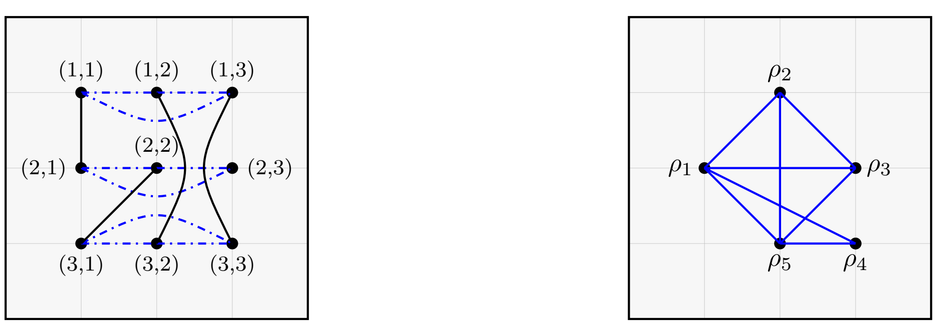
1. *We let $\bar{\rho}_G$ denote the contraction multigraph of the graph $G^{\otimes j}$ constructed on the blocks of ρ by adding an edge between two blocks ρ_1, ρ_2 of the partition ρ whenever there exist $(i, l_1) \in \rho_1$ and $(i, l_2) \in \rho_2$ such that (l_1, l_2) is an edge in G_i.*

2. *We let ρ_G be the graph constructed on the blocks of ρ by removing redundant edges in $\bar{\rho}_G$, so that at most one edge remains between any two blocks $\rho_1, \rho_2 \in \rho$.*

Example 2.4 *Consider $\rho \in \Pi([3] \times [3])$ of the form*

$$\begin{aligned}\rho &= \big\{\{(1,1),(2,1)\},\{(1,2),(3,2)\},\{(1,3),(3,3)\},\{(2,3)\},\{(2,2),(3,1)\}\big\} \\ &= \{\rho_1, \rho_2, \rho_3, \rho_4, \rho_5\},\end{aligned}$$

and let G be a triangle on the vertex set $V_G = [3]$. Figure 4 presents the multigraph $\bar{\rho}_G$ and corresponding graph ρ_G, where $\rho \in \Pi([3] \times [3])$ is non-flat and connected.



(a) Multigraph $\bar{\rho}_G$ before merging edges and vertices. (b) Graph ρ_G after merging edges and vertices.

Figure 4: Example of graph ρ_G with $j = 3$ and $k = 3$.

The contraction multigraph $\bar{\rho}_G$ constructed in Definition 2.3 is also denoted by $G^{\otimes j}$ in [FMN16, Page 104], where $G^{\otimes j}$ stands for the graph made of j components isomorphic to G.

We will use the following standard notation for the asymptotic behavior of the relative order of magnitude of two functions $f(n)$ and $g(n) > 0$ as n tends to infinity. We write

- $f(n) \gtrsim g(n)$ if $\liminf_{n\to\infty} f(n)/g(n) > 0$,
- $f(n) \ll g(n)$, or $g(n) \gg f(n)$, if $f(n) \geq 0$, $g(n) > 0$, and $\lim_{n\to\infty} f(n)/g(n) = 0$.

We close this section by recalling the following lemma, see Lemma 2.8 in [LP24a] or Proposition 6.1 in [ST24].

Lemma 2.5 *a) The cardinality of the set* $\mathrm{CNF}(j,k)$ *of connected non-flat partitions of* $[j]\times[k]$ *satisfies*

$$|\mathrm{CNF}(j,k)| \leq j!^k k!^{j-1}, \qquad j,k \geq 1.$$

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

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

with the bounds

$$((k-1)k)^{j-1}(j-1)! \leq |\mathcal{M}(j,k)| \leq ((k-1)k)^{j-1}j!, \quad j\geq 1,\ k\geq 2.$$

3 Moments of generalized U-statistics

In this section, we prove a moment identity for generalized U-statistics, see [JN91] and [Jan97, Chapter 11], by combining multiple integrals with set partitions. Let $\mathcal{M}$ and $\mathcal{S}$ be respectively a measurable space and a Borel space.

Definition 3.1 *Given* $f : \mathcal{S}^k \times \mathcal{M}^{(k-1)k/2} \to \mathbb{R}$ *a bounded measurable function,* $k \geq 2$, $\mathcal{Y} = \{Y_{i,j}\}_{1\leq i<j\leq n}$ *a sequence of independent* $\mathcal{M}$*-valued random variables with common probability distribution* $\mathbb{Q}$*, and* $\mathcal{X} = \{X_i\}_{1\leq i\leq n}$ *a sequence of independent* $\mathcal{S}$*-valued random variables with continuous distribution* μ*, we let* $S_{n,k}(f)$ *denote the generalized* U*-statistics defined as*

$$S_{n,k}(f) := \sum_{\beta\in[n]^k_{\neq}} f(X_{\beta(1)},\ldots,X_{\beta(k)},Y_{\beta(1),\beta(2)},\ldots,Y_{\beta(1),\beta(k)},Y_{\beta(2),\beta(3)},\ldots,Y_{\beta(k-1),\beta(k)}), \quad (3.1)$$

where $Y_{i,j} = Y_{j,i}$ *if* $j > i$.

Note that the function f is not required to be symmetric as in [JN91] or [Zha22]. For any $\beta = (\beta(1), \ldots, \beta(k)) \in [n]_{\neq}^{k}$, we also let

$$f(\mathbf{X}_\beta, \mathbf{Y}_\beta) = f\big(X_{\beta(1)}, \ldots, X_{\beta(k)}, Y_{\beta(1),\beta(2)}, \ldots, Y_{\beta(k-1),\beta(k)}\big)$$

for shortness of notation. Letting ρ_{K_k} denote the contraction graph of the complete graph K_k on k vertices, see Definition 2.3, for any set partition $\rho = \{b_1, \ldots, b_{|\rho|}\} \in \Pi([j] \times [k])$ we consider the function

$$\left(\bigotimes_{i=1}^{j} f\right)_\rho : \mathcal{S}^{|\rho|} \times \mathcal{M}^{e(\rho_{K_k})} \to \mathbb{R}$$

defined as

$$\left(\bigotimes_{i=1}^{j} f\right)_\rho \big(\{x_\nu\}_{1\leq \nu \leq |\rho|}, \{y_{(\nu,v)}\}_{(\nu,v)\in E_{\rho_{K_k}}}\big) := \prod_{i=1}^{j} f\big(x_1^{(i)}, \ldots, x_k^{(i)}, y_{1,2}^{(i)}, \ldots, y_{k-1,k}^{(i)}\big),$$

where $x_\ell^{(i)} := x_\nu$ if $(i,\ell) \in b_\nu$ for $1 \leq \nu \leq |\rho|$, and $y_{u,v}^{(i)} := y_{(t,s)}$ if $(i,u) \in b_t$ and $(i,v) \in b_s$, $s,t = 1, \ldots, |b|$. In Theorem 3.2 we provide a moment identity for the generalized U-statistics $S_{n,k}(f)$ using the partition diagram language of Definition 2.3.

Theorem 3.2 *Let $f : \mathcal{S}^k \times \mathcal{M}^{(k-1)k/2} \to \mathbb{R}$ be a bounded measurable function with $k \geq 2$. For any $j, n \geq 1$, we have*

$$\mathbb{E}\left[(S_{n,k}(f))^j\right] = \sum_{\substack{\rho\in\Pi([j]\times[k]) \\ \rho\wedge\pi=\widehat{0}}} \frac{n!}{(n-|\rho|)!} \int_{\mathcal{S}^{|\rho|}\times\mathcal{M}^{e(\rho_{K_k})}} \left(\bigotimes_{i=1}^{j} f\right)_\rho (\mathbf{x},\mathbf{y}) \mu^{\otimes|\rho|}(\mathrm{d}\mathbf{x}) \mathbb{Q}^{e(\rho_{K_k})}(\mathrm{d}\mathbf{y}).$$

Proof. From the definition of $\sqcap(\alpha)$, we have

$$\begin{aligned}
\mathbb{E}\left[(S_{n,k}(f))^j\right] &= \sum_{\alpha_1\in[n]_{\neq}^{k},\ldots,\alpha_j\in[n]_{\neq}^{k}} \mathbb{E}\left[\prod_{i=1}^{j} f(\mathbf{X}_{\alpha_i}, \mathbf{Y}_{\alpha_i})\right] \\
&= \sum_{\substack{\rho\in\Pi([j]\times[k]) \\ \rho\wedge\pi=\widehat{0}}} \sum_{\substack{\alpha_1\in[n]_{\neq}^{k},\ldots,\alpha_j\in[n]_{\neq}^{k} \\ \sqcap(\alpha)=\rho}} \mathbb{E}\left[\prod_{i=1}^{j} f(\mathbf{X}_{\alpha_i}, \mathbf{Y}_{\alpha_i})\right] \\
&= \sum_{\substack{\rho\in\Pi([j]\times[k]) \\ \rho\wedge\pi=\widehat{0}}} \sum_{\substack{\alpha_1\in[n]_{\neq}^{k},\ldots,\alpha_j\in[n]_{\neq}^{k} \\ \sqcap(\alpha)=\rho}} \int_{\mathcal{S}^{|\rho|}\times\mathcal{M}^{e(\rho_{K_k})}} \left(\bigotimes_{i=1}^{j} f\right)_\rho (\mathbf{x},\mathbf{y}) \mu^{\otimes|\rho|}(\mathrm{d}\mathbf{x}) \mathbb{Q}^{e(\rho_{K_k})}(\mathrm{d}\mathbf{y}) \\
&= \sum_{\substack{\rho\in\Pi([j]\times[k]) \\ \rho\wedge\pi=\widehat{0}}} C_n(\rho) \int_{\mathcal{S}^{|\rho|}\times\mathcal{M}^{e(\rho_{K_k})}} \left(\bigotimes_{i=1}^{j} f\right)_\rho (\mathbf{x},\mathbf{y}) \mu^{\otimes|\rho|}(\mathrm{d}\mathbf{x}) \mathbb{Q}^{e(\rho_{K_k})}(\mathrm{d}\mathbf{y}),
\end{aligned}$$

where

$$C_n(\rho) := \frac{n!}{(n-|\rho|)!}, \quad \rho \in \Pi([j] \times [k]),$$

denotes the count of the $\alpha = [\alpha_1, \ldots, \alpha_j]^\top \in [n]^k_{\neq} \times \cdots \times [n]^k_{\neq}$ such that $\sqcap(\alpha) = \rho$, $j, k \geq 1$, with ρ non-flat. $\square$

In particular, since $\widehat{0} \in \Pi([1] \times [k])$ is the only non-flat partition of $[1] \times [k]$, we have

$$\mathbb{E}\left[S_{n,k}(f)\right] = \frac{n!}{(n-k)!} \int_{\mathcal{S}^k \times \mathcal{M}^{(k-1)k/2}} f(\mathbf{x}, \mathbf{y}) \mu^{\otimes k}(\mathrm{d}\mathbf{x}) \mathbb{Q}^{(k-1)k/2}(\mathrm{d}\mathbf{y}). \tag{3.2}$$

Moments of standard U-statistics

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

$$\left(\bigotimes_{i=1}^j f^{(i)}\right)(x_{1,1}, \ldots, x_{1,k}, \ldots, x_{j,1}, \ldots, x_{j,k}) := \prod_{i=1}^j f^{(i)}(x_{i,1}, \ldots, x_{i,k}),$$

and for $\rho \in \Pi([j] \times [k])$ we denote by

$$\left(\bigotimes_{i=1}^j f^{(i)}\right)_\rho : (\mathbb{R}^d)^{|\rho|} \to \mathbb{R}$$

the function obtained by equating any two variables whose indexes belong to a same block of ρ. In Corollary 3.3, as a consequence of Theorem 3.2, we obtain a moment identity for standard U-statistics of order $k \geq 1$, of the form

$$S_n(f) := \sum_{\beta \in [n]^k_{\neq}} f\big(X_{\beta(1)}, \ldots, X_{\beta(k)}\big).$$

Corollary 3.3 *Let $f : (\mathbb{R}^d)^k \to \mathbb{R}$, $k \geq 1$, be a bounded, not necessarily symmetric, measurable function. For any $j, n \geq 1$, we have*

$$\mathbb{E}\left[(S_n(f))^j\right] = \sum_{\substack{\rho \in \Pi([j]\times[k]) \\ \rho \wedge \pi = \widehat{0}}} \frac{n!}{(n-|\rho|)!} \int_{(\mathbb{R}^d)^{|\rho|}} \left(\bigotimes_{i=1}^j f\right)_\rho (x_1, \ldots, x_{|\rho|}) \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_{|\rho|}).$$

Proof. Since the sequence $(X_1, \ldots, X_n)$ is i.i.d., we have

$$\begin{aligned}
\mathbb{E}\left[(S_n(f))^j\right] &= \sum_{\alpha_1 \in [n]^k_{\neq}, \ldots, \alpha_j \in [n]^k_{\neq}} \mathbb{E}\left[\prod_{i=1}^j f\big(X_{\alpha_i(1)}, \ldots, X_{\alpha_i(k)}\big)\right] \\
&= \sum_{\substack{\rho \in \Pi([j]\times[k]) \\ \rho \wedge \pi = \widehat{0}}} C_n(\rho) \int_{(\mathbb{R}^d)^{|\rho|}} \left(\bigotimes_{i=1}^j f\right)_\rho (x_1, \ldots, x_{|\rho|}) \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_{|\rho|}).
\end{aligned}$$

$\square$

Remark 3.4 *For the V-statistics*

$$V_{n,k}(f) := \sum_{\beta\in[n]^k} f\big(X_{\beta(1)},\dots,X_{\beta(k)}\big)$$

with possible repeated indices, for $j,n\geq 1$, we have the similar moment identity

$$\mathbb{E}\left[(V_{n,k}(f))^j\right] = \sum_{\rho\in\Pi([j]\times[k])} \frac{n!}{(n-|\rho|)!}\int_{(\mathbb{R}^d)^{|\rho|}} \left(\bigotimes_{i=1}^j f\right)_\rho (x_1,\dots,x_{|\rho|})\mu(\mathrm{d}x_1)\cdots\mu(\mathrm{d}x_{|\rho|}). \quad (3.3)$$

The proof of (3.3) *is almost the same, except for the removal of the non-flat restriction, i.e.*

$$\begin{aligned}\mathbb{E}\left[(V_{n,k}(f))^j\right] &= \sum_{\alpha_1,\dots,\alpha_j\in[n]^k} \mathbb{E}\left[\prod_{i=1}^j f\big(X_{\alpha_i(1)},\dots,X_{\alpha_i(k)}\big)\right]\\ &= \sum_{\rho\in\Pi([j]\times[k])} \frac{n!}{(n-|\rho|)!}\int_{(\mathbb{R}^d)^{|\rho|}} \left(\bigotimes_{i=1}^j f\right)_\rho (x_1,\dots,x_{|\rho|})\mu(\mathrm{d}x_1)\cdots\mu(\mathrm{d}x_{|\rho|}).\end{aligned}$$

4 Cumulant bounds for generalized U-statistics

To study the asymptotic behaviour of U-statistics and generalized U-statistics, a common practice is to use Hoeffding decompositions [Hoe61], [DM83]. Through the orthogonal decomposition, one finds the asymptotic distribution of U-statistics is determined by the "smallest" component appearing in its decomposition, see [JN91, Lemma 2] and also [Jan97, Theorem 11.3]. In what follows, we consider the case where the asymptotic distribution of $S_{n,k}(f)$ is normal. Assumption 4.1 will be needed for the derivation of cumulant estimates for the generalized U-statistics $S_{n,k}(f)$ in Theorem 4.1.

Assumption 4.1 *Let $f:\mathcal{S}^k\times\mathcal{M}^{(k-1)k/2}\to\mathbb{R}$ be a bounded measurable function with $k\geq 2$. We assume that*

$$\operatorname{Var}\left[\sum_{\ell=1}^k f_{(\ell)}(X_1)\right] > 0. \quad (4.1)$$

where

$$f_{(i)}(x) := \int_{\mathcal{S}^{k-1}\times\mathcal{M}^{(k-1)k/2}} f(\mathbf{x},\mathbf{y})\mu^{\otimes(k-1)}\{\mathrm{d}x_1\cdots\mathrm{d}x_{i-1}\mathrm{d}x_{i+1}\cdots\mathrm{d}x_k\}\mathbb{Q}^{(k-1)k/2}\{\mathrm{d}\mathbf{y}\}$$

for $i=1,\dots,k$, and $\mathbf{x}=(x_1,\dots,x_{i-1},x,x_{i+1},\dots,x_k)$.

Assumption 4.1 amounts to saying that the principal degree of f equals 1, see [JN91]. In most of the existing literature, including [JN91], [KR21], [Zha22], the asymptotic normality of generalized U-statistics relies on orthogonal decomposition on certain L_2 spaces. However, in practice the orthogonal decomposition of count statistics can be intractable in general.

Theorem 4.1 *Let $f : \mathcal{S}^k \times \mathcal{M}^{(k-1)k/2} \to \mathbb{R}$ be a measurable function, with $k \geq 2$. Then, for $n \geq 1$ the j-th cumulant of the generalized U-statistics $S_{n,k}(f)$ satisfies the bound*

$$|\kappa_j(S_{n,k}(f))| \leq n^{1+(k-1)j} \|f\|_\infty^j j^{j-1} (j!)^k (k!)^{j-1}, \qquad j \geq 1. \tag{4.2}$$

In addition, if the function f satisfies Assumption 4.1, then we have

$$\kappa_2(S_{n,k}(f)) \geq \frac{Cn!}{(n-2k+1)!}, \quad n \geq N(f,k), \tag{4.3}$$

where $C(f,k) > 0$ and $N(f,k) \geq 1$ depend only on f and on $k \geq 2$.

Proof. From Property C in [MM91, Page 29], we know that if $\alpha = [\alpha_1, \ldots, \alpha_j]^\top \in [n]_{\neq}^k \times \cdots \times [n]_{\neq}^k$ is disconnected, i.e. if there exists a partition $\{A, B\}$ of $[j]$ such that

$$\{\alpha_i(\ell)\}_{(i,\ell)\in A\times[k]} \cap \{\alpha_i(\ell)\}_{(i,\ell)\in B\times[k]} = \emptyset,$$

then the joint cumulant $\kappa(f(\mathbf{X}_{\alpha_1}, \mathbf{Y}_{\alpha_1}), \ldots, f(\mathbf{X}_{\alpha_j}, \mathbf{Y}_{\alpha_j}))$ vanishes. Hence, we have

$$\begin{aligned}
\kappa_j(S_{n,k}(f)) &= \sum_{\alpha_1,\ldots,\alpha_j \in [n]_{\neq}^k} \kappa(f(\mathbf{X}_{\alpha_1}, \mathbf{Y}_{\alpha_1}), \ldots, f(\mathbf{X}_{\alpha_j}, \mathbf{Y}_{\alpha_j})) \\
&= \sum_{\substack{\alpha_1 \in [n]_{\neq}^k, \ldots, \alpha_j \in [n]_{\neq}^k \\ \text{connected}}} \kappa(f(\mathbf{X}_{\alpha_1}, \mathbf{Y}_{\alpha_1}), \ldots, f(\mathbf{X}_{\alpha_j}, \mathbf{Y}_{\alpha_j})) \\
&= \sum_{\rho \in \mathrm{CNF}(j,k)} \sum_{\substack{\alpha = [\alpha_1, \ldots, \alpha_j]^\top \in ([n]_{\neq}^k)^j \\ \sqcap(\alpha) = \rho}} \kappa(f(\mathbf{X}_{\alpha_1}, \mathbf{Y}_{\alpha_1}), \ldots, f(\mathbf{X}_{\alpha_j}, \mathbf{Y}_{\alpha_j})).
\end{aligned} \tag{4.4}$$

By the cumulant-moment relation (A.1) we have, for any $[\alpha_1, \ldots, \alpha_j]^\top \in [n]_{\neq}^k \times \cdots \times [n]_{\neq}^k$,

$$\begin{aligned}
\left|\kappa\big(f(\mathbf{X}_{\alpha_1}, \mathbf{Y}_{\alpha_1}), \ldots, f(\mathbf{X}_{\alpha_j}, \mathbf{Y}_{\alpha_j})\big)\right| &\leq \sum_{\sigma = \{b_1, \ldots, b_l\} \in \Pi([j])} (l-1)! \prod_{i=1}^{l} \left| \mathbb{E}\left[\prod_{\ell \in b_i} f(\mathbf{X}_{\alpha_\ell}, \mathbf{Y}_{\alpha_\ell}) \right] \right| \\
&\leq \sum_{\sigma = \{b_1, \ldots, b_l\} \in \Pi([j])} (l-1)! \|f\|_\infty^j \\
&= \|f\|_\infty^j \sum_{l=1}^{j} (l-1)! S(j,l) \\
&\leq \|f\|_\infty^j \frac{1}{j} \sum_{l=1}^{j} \frac{j!}{(j-l)!} S(j,l) \\
&= \|f\|_\infty^j j^{j-1},
\end{aligned} \tag{4.5}$$

where $S(j,l)$ is the Stirling number of the second kind. Therefore, from (4.4) we obtain

$$\kappa_j(S_{n,k}(f)) \leq \sum_{\rho \in \mathrm{CNF}(j,k)} \frac{n!}{(n-|\rho|)!} \|f\|_\infty^j j^{j-1}$$

$$= \sum_{r=k}^{1+(k-1)j} \frac{n!}{(n-r)!} \sum_{\substack{\rho\in\mathrm{CNF}(j,k)\\|\rho|=r}} \|f\|_\infty^j j^{j-1}$$
$$\le n^{1+(k-1)j}\|f\|_\infty^j j^{j-1}|\mathrm{CNF}(j,k)|,$$

which yields (4.2) from Lemma 2.5-(a). Letting $j=2$ in (4.4), we have

$$\begin{aligned}
\kappa_2(S_{n,k}(f)) &= \sum_{\rho\in\mathrm{CNF}(2,k)} \sum_{\substack{\alpha=[\alpha_1,\alpha_2]^\top\in([n]^k_{\neq})^2\\ \sqcap(\alpha)=\rho}} \kappa(f(\mathbf{X}_{\alpha_1},\mathbf{Y}_{\alpha_1}), f(\mathbf{X}_{\alpha_2},\mathbf{Y}_{\alpha_2}))\\
&= \sum_{\rho\in\mathrm{CNF}(2,k)} \sum_{\substack{\alpha=[\alpha_1,\alpha_2]^\top\in([n]^k_{\neq})^2\\ \sqcap(\alpha)=\rho}} \mathrm{Cov}(f(\mathbf{X}_{\alpha_1},\mathbf{Y}_{\alpha_1}), f(\mathbf{X}_{\alpha_2},\mathbf{Y}_{\alpha_2}))\\
&= R + \sum_{\substack{\rho\in\mathrm{CNF}(2,k)\\|\rho|=2k-1}} \sum_{\substack{\alpha=[\alpha_1,\alpha_2]^\top\in([n]^k_{\neq})^2\\ \sqcap(\alpha)=\rho}} \mathrm{Cov}(f(\mathbf{X}_{\alpha_1},\mathbf{Y}_{\alpha_1}), f(\mathbf{X}_{\alpha_2},\mathbf{Y}_{\alpha_2})),
\end{aligned} \tag{4.6}$$

where

$$R := \sum_{\substack{\rho\in\mathrm{CNF}(2,k)\\|\rho|\le 2k-2}} \sum_{\substack{\alpha=[\alpha_1,\alpha_2]^\top\in([n]^k_{\neq})^2\\ \sqcap(\alpha)=\rho}} \mathrm{Cov}(f(\mathbf{X}_{\alpha_1},\mathbf{Y}_{\alpha_1}), f(\mathbf{X}_{\alpha_2},\mathbf{Y}_{\alpha_2})).$$

Because $\{X_1,\dots,X_n\}$ and $\{Y_{i,j}\}_{1\le i<j\le n}$ are both i.i.d. random elements, we have, for any $\alpha=[\alpha_1,\alpha_2]^\top\in([n]^k_{\neq})^2$, $\beta=[\beta_1,\beta_2]^\top\in([n]^k_{\neq})^2$

$$\mathrm{Cov}(f(\mathbf{X}_{\alpha_1},\mathbf{Y}_{\alpha_1}), f(\mathbf{X}_{\alpha_2},\mathbf{Y}_{\alpha_2})) = \mathrm{Cov}(f(\mathbf{X}_{\beta_1},\mathbf{Y}_{\beta_1}), f(\mathbf{X}_{\beta_2},\mathbf{Y}_{\beta_2})),$$

if $\sqcap(\alpha)=\sqcap(\beta)$. Let $\sigma\in\mathrm{CNF}(2,k)$ with $|\sigma|=2k-1$ and $\{(1,s),(2,t)\}\in\sigma$, $1\le s\neq t\le k$. Taking $\alpha=[\alpha_1,\alpha_2]^\top\in([n]^k_{\neq})^2$ such that $\sqcap(\alpha)=\sigma$ we have $\alpha_1(s)=\alpha_2(t)$, hence

$$\begin{aligned}
&\mathrm{Cov}(f(\mathbf{X}_{\alpha_1},\mathbf{Y}_{\alpha_1}), f(\mathbf{X}_{\alpha_2},\mathbf{Y}_{\alpha_2}))\\
&\quad= \mathbb{E}\left[f(\mathbf{X}_{\alpha_1},\mathbf{Y}_{\alpha_1})f(\mathbf{X}_{\alpha_2},\mathbf{Y}_{\alpha_2})\right] - \mathbb{E}\left[f(\mathbf{X}_{\alpha_1},\mathbf{Y}_{\alpha_1})\right] - \mathbb{E}\left[f(\mathbf{X}_{\alpha_2},\mathbf{Y}_{\alpha_2})\right]\\
&\quad= \mathbb{E}\left[\mathbb{E}[f(\mathbf{X}_{\alpha_1},\mathbf{Y}_{\alpha_1})f(\mathbf{X}_{\alpha_2},\mathbf{Y}_{\alpha_2})|X_{\alpha_1(t)}]\right] - \mathbb{E}\left[\mathbb{E}[f(\mathbf{X}_{\alpha_1},\mathbf{Y}_{\alpha_1})|X_{\alpha_1(s)}]\right]\mathbb{E}\left[\mathbb{E}[f(\mathbf{X}_{\alpha_2},\mathbf{Y}_{\alpha_2})|X_{\alpha_2(t)}]\right]\\
&\quad= \mathbb{E}\left[f_{(s)}(X_{\alpha_1(s)})f_{(t)}(X_{\alpha_1(t)})\right] - \mathbb{E}\left[f_{(s)}(X_{\alpha_1(s)})\right]\mathbb{E}\left[f_{(t)}(X_{\alpha_1(t)})\right]\\
&\quad= \mathbb{E}\left[f_{(s)}(X_1)f_{(t)}(X_1)\right] - \mathbb{E}\left[f_{(s)}(X_1)\right]\mathbb{E}\left[f_{(t)}(X_1)\right]\\
&\quad= \mathrm{Cov}(f_{(s)}(X_1), f_{(t)}(X_1))\\
&\quad=: r_\sigma.
\end{aligned}$$

Therefore, the second term in (4.6) becomes

$$\sum_{\substack{\rho\in\mathrm{CNF}(2,k)\\|\rho|=2k-1}} \sum_{\substack{\alpha=[\alpha_1,\alpha_2]^\top\in([n]^k_{\neq})^2\\ \sqcap(\alpha)=\rho}} \mathrm{Cov}(f(\mathbf{X}_{\alpha_1},\mathbf{Y}_{\alpha_1}), f(\mathbf{X}_{\alpha_2},\mathbf{Y}_{\alpha_2}))$$

$$
\begin{aligned}
&= \sum_{\substack{\rho\in\mathrm{CNF}(2,k)\\|\rho|=2k-1}} \sum_{\substack{\alpha=[\alpha_1,\alpha_2]^\top\in([n]^k_{\neq})^2\\ \sqcap(\alpha)=\rho}} r_\rho\\
&= \sum_{\substack{\rho\in\mathrm{CNF}(2,k)\\|\rho|=2k-1}} \frac{n!}{(n-2k+1)!} r_\rho\\
&= \frac{n!}{(n-2k+1)!} \sum_{\substack{\rho\in\mathrm{CNF}(2,k)\\|\rho|=2k-1}} r_\rho\\
&= \frac{n!}{(n-2k+1)!} \sum_{1\le s,t\le k} \mathrm{Cov}(f_{(s)}(X_1), f_{(t)}(X_1))\\
&= \frac{n!}{(n-2k+1)!} \mathrm{Var}\left[\sum_{\ell=1}^{k} f_{(\ell)}(X_1)\right]. \qquad (4.7)
\end{aligned}
$$

On the other hand, from (4.5), we have

$$
\begin{aligned}
|R| &\le \sum_{\substack{\rho\in\mathrm{CNF}(2,k)\\|\rho|\le 2k-2}} \sum_{\substack{\alpha=[\alpha_1,\alpha_2]^\top\in([n]^k_{\neq})^2\\ \sqcap(\alpha)=\rho}} |\mathrm{Cov}(f(\mathbf{X}_{\alpha_1},\mathbf{Y}_{\alpha_1}), f(\mathbf{X}_{\alpha_2},\mathbf{Y}_{\alpha_2}))|\\
&\le \sum_{\substack{\rho\in\mathrm{CNF}(2,k)\\|\rho|\le 2k-2}} \frac{n!}{(n-|\rho|)!} 2\|f\|_\infty^2\\
&\le 2\|f\|_\infty^2 n^{2k-2} |\mathrm{CNF}(2,k)|\\
&\le 2\|f\|_\infty^2 n^{2k-2} 2^k k!. \qquad (4.8)
\end{aligned}
$$

Finally, we bound $\mathrm{Var}[S_{n,k}(f)]$ from below using (4.6), (4.7) and (4.8), as

$$
\begin{aligned}
\kappa_2(S_{n,k}(f)) &= \sum_{\substack{\rho\in\mathrm{CNF}(2,k)\\|\rho|=2k-1}} \sum_{\substack{\alpha=[\alpha_1,\alpha_2]^\top\in([n]^k_{\neq})^2\\ \sqcap(\alpha)=\rho}} \mathrm{Cov}(f(\mathbf{X}_{\alpha_1},\mathbf{Y}_{\alpha_1}), f(\mathbf{X}_{\alpha_2},\mathbf{Y}_{\alpha_2})) + R\\
&\ge \frac{n!}{(n-2k+1)!} \mathrm{Var}\left[\sum_{\ell=1}^{k} f_{(\ell)}(X_1)\right] - 2\|f\|_\infty^2 n^{2k-2} 2^k k!\\
&= \frac{n!}{(n-2k+1)!} \left(\mathrm{Var}\left[\sum_{\ell=1}^{k} f_{(\ell)}(X_1)\right] - \frac{2\|f\|_\infty^2 n^{2k-2} 2^k k!}{n!/(n-2k+1)!}\right)\\
&\ge \frac{n!}{(n-2k+1)!} \left(\mathrm{Var}\left[\sum_{\ell=1}^{k} f_{(\ell)}(X_1)\right] - \frac{2\|f\|_\infty^2 n^{2k-2} 2^k k!}{(n/2)^{2k-1}}\right),
\end{aligned}
$$

which yields (4.3) by choosing

$$
C(f,k) := \frac{1}{2}\mathrm{Var}\left[\sum_{\ell=1}^{k} f_{(\ell)}(X_1)\right],
$$

for

$$n \geq N(f,k) := \frac{2^{3k+1} \|f\|_\infty^2 k!}{\operatorname{Var}\left[\sum_{\ell=1}^k f_{(\ell)}(X_1)\right]}.$$

$\square$

As a consequence of Theorem 4.1, we have the following result.

Corollary 4.2 *Let $f : \mathcal{S}^k \times \mathcal{M}^{(k-1)k/2} \to \mathbb{R}$ be a measurable function with $k \geq 2$, such that f satisfies Assumption 4.1. Then, for $n \geq 4(k-1)$, the j-th cumulant of the normalized U-statistics*

$$\overline{S}_{n,k}(f) := \frac{S_{n,k}(f) - \kappa_1(S_{n,k}(f))}{\sqrt{\kappa_2(S_{n,k}(f))}}$$

satisfies the bound

$$\kappa_j\left(\overline{S}_{n,k}(f)\right) \leq \frac{(j!)^{k+1}}{\left(n\widetilde{C}(f,k)\right)^{j/2-1}},$$

$j \geq 3$, where $\widetilde{C}(f,k) > 0$ depends only on f and on $k \geq 2$.

Proof. Combining the inequalities

$$\frac{n!}{(n-2k+1)!} \geq (n-2k+2)^{2k-1} = \left(1 - \frac{2k-2}{n}\right)^{2k-1} n^{2k-1} \geq \left(\frac{n}{2}\right)^{2k-1}, \tag{4.9}$$

$n \geq 4(k-1)$, and

$$j^{j-1} < \frac{(j-1)!}{\sqrt{2\pi j}} e^j < e^j j!, \quad j \geq 1, \tag{4.10}$$

with the bounds (4.2) and (4.3) for $j \geq 3$, we have

$$\begin{aligned}
\kappa_j\left(\overline{S}_{n,k}\right) &= \frac{\kappa_j\left(S_{n,k}(f)\right)}{\kappa_2(S_{n,k}(f))^{j/2}} \\
&\leq \|f\|_\infty^j j^{j-1} (j!)^k (k!)^{j-1} \frac{n^{1+(k-1)j}}{(n!C(f,k)/(n-2k+1)!)^{j/2}} \\
&\leq \|f\|_\infty^j j^{j-1} (j!)^k (k!)^{j-1} \frac{n^{1+(k-1)j}}{(C(f,k)(n/2)^{2k-1})^{j/2}} \\
&= \|f\|_\infty^j \frac{j^{j-1}(j!)^k}{n^{(j-2)/2}} \frac{(k!)^{j-1}}{C(f,k)^{j/2}} \\
&\leq \|f\|_\infty^j \frac{(j!)^{k+1}}{n^{(j-2)/2}} \frac{(k!)^{j-1} e^j}{C(f,k)^{j/2}} \\
&\leq \frac{(j!)^{k+1}}{\left(n\widetilde{C}(f,k)\right)^{j/2-1}},
\end{aligned}$$

where $\widetilde{C}(f,k) > 0$ depends only on f and on $k \geq 2$. $\square$

From Corollary 4.2, we check that the cumulants of the normalized U-statistics $\overline{S}_{n,k}$ satisfy the Statulevičius growth condition (A.2) with $\gamma := k$ and

$$\Delta_n := \left(n\widetilde{C}(f,k)\right)^{1/2},$$

where $\widetilde{C}(f,k) > 0$ depends only on f and on $k \geq 2$. Therefore, from Proposition A.1-i) we have the following results.

Corollary 4.3 *(Kolmogorov bound). Let $f : \mathcal{S}^k \times \mathcal{M}^{(k-1)k/2} \to \mathbb{R}$ be a measurable function satisfying Assumption 4.1, with $k \geq 2$. For $n \geq 4(k-1)$, we have*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\overline{S}_{n,k}(f) \leq x\right) - \Phi(x) \right| \leq \frac{C(f,k)}{n^{1/(2+4k)}},$$

where $C(f,k) > 0$ depends only on f and on $k \geq 2$.

By Corollary 4.2 and Proposition A.1-ii), we also have the following result.

Corollary 4.4 *(Moderate deviation principle). Let $f : \mathcal{S}^k \times \mathcal{M}^{(k-1)k/2} \to \mathbb{R}$ be a measurable function satisfying Assumption 4.1, with $k \geq 2$. Let $(a_n)_{n\geq 1}$ be a sequence of real numbers tending to infinity, and such that $a_n \ll n^{1/(2+4k)}$. Then, $\left(a_n^{-1}\overline{S}_{n,k}(f)\right)_{n\geq 1}$ satisfies a moderate deviation principle with speed a_n^2 and rate function $x^2/2$.*

5 Upper bounds on subgraph count cumulants

From this section, we focus on the subgraph count in the binomial RCM $G_H(\mathcal{X}_n)$. Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a set of i.i.d. random points on the carrier space $\mathcal{S} = \mathbb{R}^d$ with a common continuous distribution μ. A symmetric measurable function $H : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ is called a connection function. We consider the binomial random-connection model generated as follows: given $\mathcal{X}_n$, we connect each pair of points/nodes $X_i, X_j \in \mathcal{X}_n$, $1 \leq i \neq j \leq n$, independently with probability $p_n H(X_i, X_j)$, $0 < p_n < 1$. The resulting random graph is denoted by $G_H(\mathcal{X}_n)$.

Given $G = (V_G, E_G)$ a connected graph with $v(G) = k \geq 2$ vertices, let

$$I_\beta := \prod_{(\imath,\jmath)\in E_G} \mathbf{1}_{\{X_{\beta(\imath)} \sim X_{\beta(\jmath)}\}}, \quad \beta \in [n]^k_{\neq},$$

where $X_\imath \sim X_\jmath$ indicates there is an edge between $X_\imath$ and $X_\jmath$, i.e. $I_\beta = 1$ if and only if the graph with vertices $X_{\beta(1)}, \ldots, X_{\beta(k)}$ in the binomial RCM $G_H(\mathcal{X}_n)$ is isomorphic to G.

We let

$$N_G := \frac{1}{a(G)} \sum_{\beta \in [n]^k_{\neq}} I_\beta \tag{5.1}$$

denote the number of injective subgraphs of $G_H(\mathcal{X}_n)$ isomorphic to G, where $X_i \sim X_j$ indicates that there is an edge between X_i and X_j.

In order to incorporate with generalized U-statistics in Section 3, we can also write the subgraph count N_G as

$$N_G = S_{n,k}(f) = \sum_{\beta \in [n]^k_{\neq}} f\big(X_{\beta(1)}, \ldots, X_{\beta(k)}, Y_{\beta(1),\beta(2)}, \ldots, Y_{\beta(1),\beta(k)}, Y_{\beta(2),\beta(3)}, \ldots, Y_{\beta(k-1),\beta(k)}\big),$$

with $f : (\mathbb{R}^d)^k \times [0,1]^{(k-1)k/2} \to \mathbb{R}$ given by

$$f(x_1, \ldots, x_k, y_{1,2}, \ldots, y_{(k-1),k}) := \frac{1}{a(G)} \prod_{(\imath,\jmath) \in E_G} \mathbf{1}_{\{y_{\imath,\jmath} \leq p_n H(x_\imath, x_\jmath)\}}, \tag{5.2}$$

where $\{Y_{i,j}\}_{1 \leq i < j \leq n}$ are i.i.d. uniform random variables on $\mathcal{M} = [0,1]$.

Next, using dependency graphs and the convex analysis of planar diagrams, we consider the cumulant growth of the subgraph count N_G when $p_n = o(1)$. Before we proceed, we need to introduce some notation. Since the random variables $X_1, \ldots, X_n$ are i.i.d. and edges are added independently according to

$$\mathbf{1}_{\{X_i \sim X_j\}} = \mathbf{1}_{\{Y_{i,j} \leq p_n H(X_i, X_j)\}}, \quad 1 \leq i < j \leq n,$$

the sets of random vectors

$$\{(I_{\alpha_1}, \ldots, I_{\alpha_j}) : \sqcap\big([\alpha_1, \ldots, \alpha_j]^\top\big) = \rho\}$$

are identically distributed and have the same dependency structure for every $\rho \in \Pi([j] \times [k])$, which justifies the following definition.

Definition 5.1 *Let $j, k \geq 1$. Given $\rho \in \Pi([j] \times [k])$, we let*

$$\kappa(\mathbf{I}_\rho) := \kappa(I_{\alpha_1}, \ldots, I_{\alpha_j}),$$

for any element $[\alpha_1, \ldots, \alpha_j]^\top$ of $[n]^k_{\neq} \times \cdots \times [n]^k_{\neq}$ such that $\sqcap(\alpha) = \rho$.

As a consequence of (3.2), we have

$$\kappa_1(N_G) = \frac{n!}{(n-k)!} \frac{p_n^{e(G)}}{a(G)} \int_{(\mathbb{R}^d)^k} \prod_{(\imath,\jmath) \in E_G} H(x_\imath, x_\jmath) \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_k), \tag{5.3}$$

while the growth of higher order cumulants can be controlled as follows. See Example 6.19 in [JŁR00] and Proposition 10.1.2 and § 10.3.3 of [FMN16] for related cumulant bounds in the Erdős-Rényi model.

Theorem 5.2 *Let $G = (V_G, E_G)$ be a connected graph with $v(G) = k \geq 2$ vertices. We have the cumulant bound*

$$|\kappa_j(N_G)| \leq \frac{j^{j-1}}{a(G)^j} \sum_{r=k}^{1+(k-1)j} |\mathcal{C}(j,k,r)| n^r p_n^{d(j,k,r)}, \qquad j \geq 2, \tag{5.4}$$

where $\mathcal{C}(j,k,r)$ represents the collection of all partitions ρ on $[j] \times [k]$ connected, non-flat and having precisely r blocks, and

$$d(j,k,r) := \min \big\{ e(\rho_G) \ : \ \rho \in \mathrm{CNF}(j,k), \ |\rho| = r \big\}. \tag{5.5}$$

Proof. The random variables

$$I_{\beta_1} := \prod_{(\imath,\jmath)\in E_G} \mathbf{1}_{\{X_{\beta_1(\imath)} \sim X_{\beta_1(\jmath)}\}} \quad \text{and} \quad I_{\beta_2} := \prod_{(\imath,\jmath)\in E_G} \mathbf{1}_{\{X_{\beta_2(\imath)} \sim X_{\beta_2(\jmath)}\}}, \quad \beta_1, \beta_2 \in [n]^k_{\neq},$$

are independent if

$$\big\{\beta_1(i) : i \in [k]\big\} \cap \big\{\beta_2(i) : i \in [k]\big\} = \emptyset.$$

Hence, by Property C in [MM91, Page 29] we have

$$\begin{aligned}
\kappa_j(N_G) &= \frac{1}{a(G)^j} \sum_{\alpha_1,\ldots,\alpha_j \in [n]^k_{\neq}} \kappa(I_{\alpha_1}, \ldots, I_{\alpha_j}) \\
&= \frac{1}{a(G)^j} \sum_{\substack{\alpha_1,\ldots,\alpha_j \in [n]^k_{\neq} \\ \text{connected}}} \kappa(I_{\alpha_1}, \ldots, I_{\alpha_j}) \\
&= \frac{1}{a(G)^j} \sum_{[\alpha_1,\ldots,\alpha_j]^\top \in \Lambda_j(n,k)} \kappa(I_{\alpha_1}, \ldots, I_{\alpha_j}), \qquad j \geq 2,
\end{aligned}$$

where

$$\begin{aligned}
\Lambda_j(n,k) &:= \big\{ \alpha = [\alpha_1, \ldots, \alpha_j]^\top \in [n]^k_{\neq} \times \cdots \times [n]^k_{\neq} \text{ connected} \big\} \\
&= \big\{ \alpha = [\alpha_1, \ldots, \alpha_j]^\top \in [n]^k_{\neq} \times \cdots \times [n]^k_{\neq} : \sqcap(\alpha) \in \mathrm{CNF}(j,k) \big\}.
\end{aligned}$$

Now, we have

$$\kappa_j(N_G) = \frac{1}{a(G)^j} \sum_{\rho \in \mathrm{CNF}(j,k)} \sum_{\substack{[\alpha_1,\ldots,\alpha_j]^\top \in \Lambda_j(n,k) \\ \sqcap([\alpha_1,\ldots,\alpha_j]^\top) = \rho}} \kappa(I_{\alpha_1}, \ldots, I_{\alpha_j})$$

$$
\begin{aligned}
&= \frac{1}{a(G)^j} \sum_{\rho \in \mathrm{CNF}(j,k)} C_n(\rho)\kappa(\mathbf{I}_\rho) \\
&= \frac{1}{a(G)^j} \sum_{\rho \in \mathrm{CNF}(j,k)} \frac{n!}{(n-|\rho|)!}\kappa(\mathbf{I}_\rho) \\
&= \frac{1}{a(G)^j} \sum_{r=k}^{1+(k-1)j} \frac{n!}{(n-r)!} \sum_{\substack{\rho \in \mathrm{CNF}(j,k) \\ |\rho|=r}} \kappa(\mathbf{I}_\rho), \qquad (5.6)
\end{aligned}
$$

hence

$$
|\kappa_j(N_G)| \leq \frac{1}{a(G)^j} \sum_{r=k}^{1+(k-1)j} \frac{n!}{(n-r)!} |\mathcal{C}(j,k,r)| \max_{\rho \in \mathcal{C}(j,k,r)} |\kappa(\mathbf{I}_\rho)|. \qquad (5.7)
$$

Due to the cumulant-moment relation (A.1), we have, for each $\alpha = [\alpha_1, \ldots, \alpha_j]^\top \in \Lambda_j(n,k)$,

$$
\begin{aligned}
\kappa(I_{\alpha_1}, \ldots, I_{\alpha_j}) &= \sum_{s \geq 1} \sum_{\sigma = \{b_1, \ldots, b_s\} \in \Pi([j])} (-1)^{s-1}(s-1)! \mathbb{E}\left[\prod_{i \in b_1} I_{\alpha_i}\right] \times \cdots \times \mathbb{E}\left[\prod_{i \in b_s} I_{\alpha_i}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{j} I_{\alpha_i}\right] + \sum_{s \geq 2} \sum_{\sigma = \{b_1, \ldots, b_s\} \in \Pi([j])} (-1)^{s-1}(s-1)! \mathbb{E}\left[\prod_{i \in b_1} I_{\alpha_i}\right] \times \cdots \times \mathbb{E}\left[\prod_{i \in b_s} I_{\alpha_i}\right].
\end{aligned}
$$

Moreover, denoting by $\sqcap(\alpha)_G$ the connected graph built on $\sqcap(\alpha) \in \Pi([j] \times [k])$ in Definition 2.3, we have

$$
\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^{j} I_{\alpha_i}\right] &= \mathbb{E}\left[\prod_{i=1}^{j} \left(\prod_{(\imath,\jmath) \in E(G)} \mathbf{1}_{\{X_{\alpha_i(\imath)} \sim X_{\alpha_i(\jmath)}\}} \right)\right] \\
&= p_n^{e(\sqcap(\alpha)_G)} \int_{(\mathbb{R}^d)^{|\sqcap(\alpha)|}} \prod_{(\imath,\ell) \in E(\sqcap(\alpha)_G)} H(x_\imath, x_\ell) \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_{|\sqcap(\alpha)|}) \\
&\leq p_n^{e(\sqcap(\alpha)_G)},
\end{aligned}
$$

where the last inequality is due to the facts that $0 \leq H \leq 1$ and μ is a probability measure on $\mathbb{R}^d$. For each $\rho \in \Pi([j] \times [k])$ and $b \subset [j]$, we denote $\rho_{|b} := \rho_{|\{\pi_i \,:\, i \in b\}}$ the partition of $\cup_{i \in b} \pi_i$ obtained restricting ρ on $\cup_{i \in b} \pi_i$, i.e.

$$
\rho_{|b} = \{a \cap (\cup_{i \in b} \pi_i) \,:\, a \in \rho\}.
$$

Therefore, for any $\sigma = \{b_1, \ldots, b_r\} \in \Pi([j])$ with $r \geq 2$ we have, letting $\rho := \sqcap(a)$,

$$
\prod_{i=1}^{r} \mathbb{E}\left[\prod_{j \in b_i} I_{\alpha_j}\right] = \prod_{i=1}^{r} \mathbb{E}\left[\prod_{\ell \in b_i} \left(\prod_{(\imath,\jmath) \in E(G)} \mathbf{1}_{\{X_{\alpha_\ell(\imath)} \sim X_{\alpha_\ell(\jmath)}\}} \right)\right]
$$

$$
\begin{aligned}
&= \prod_{i=1}^{r} p_n^{e((\rho_{|b_i})_G)} \int_{(\mathbb{R}^d)^{|\rho_{|b_i}|}} \prod_{(\imath,\ell)\in E((\rho_{|b_i})_G)} H(x_\imath, x_\ell)\mu(\mathrm{d}x_1)\cdots\mu(\mathrm{d}x_{|\rho_{|b_i}|}) \\
&\le p_n^{\sum_{i=1}^r e((\rho_{|b_i})_G)} \\
&\le p_n^{e(\rho_G)} \\
&= p_n^{e(\sqcap(\alpha)_G)}.
\end{aligned}
$$

Next, we have

$$
\begin{aligned}
\max_{\rho\in\mathcal{C}(j,k,r)} |\kappa(\mathbf{I}_\rho)| &\le p_n^{d(j,k,r)} \sum_{s=1}^{j} \sum_{\sigma=\{b_1,\dots,b_s\}\in\Pi([j])} (s-1)! \\
&= p_n^{d(j,k,r)} \sum_{s=1}^{j} (s-1)! S(j,s) \\
&\le \frac{p_n^{d(j,k,r)}}{j} \sum_{s=1}^{j} (j)_s S(j,s) \\
&\le p_n^{d(j,k,r)} j^{j-1},
\end{aligned}
$$

where $S(j,r)$ is the Stirling number of the second kind. Therefore, together with (5.7), we obtain

$$
\begin{aligned}
|\kappa_j(N_G)| &\le \frac{1}{a(G)^j} \sum_{r=k}^{1+(k-1)j} \frac{n!}{(n-r)!} |\mathcal{C}(j,k,r)| p_n^{d(j,k,r)} j^{j-1} \\
&\le \frac{j^{j-1}}{a(G)^j} \sum_{r=k}^{1+(k-1)j} |\mathcal{C}(j,k,r)| n^r p_n^{d(j,k,r)},
\end{aligned}
$$

which proves (5.4). □

6 Variance lower bounds for subgraph counts

To prove Theorem 7.5, we shall derive a variance lower bound for the subgraph count N_G in the binomial RCM. For this purpose, we need to introduce some notation. Recall f from (5.2), we further denote, for $1 \le \ell \le k$,

$$
\begin{aligned}
f_{(\ell)}(x) &:= \frac{1}{a(G)} \mathbb{E}\left[\prod_{(\imath,\jmath)\in E_G} \mathbf{1}_{\{Y_{\imath,\jmath} \le p_n H(X_\imath, X_\jmath)\}} \middle| X_\ell = x \right] \\
&= \int_{(\mathbb{R}^d)^{k-1}\times[0,1]^{(k-1)k/2}} f(\mathbf{x}, \mathbf{y}) \mu^{\otimes(k-1)}(\mathrm{d}x_1, \dots, \mathrm{d}x_{\ell-1}, \mathrm{d}x_{\ell+1}, \dots, \mathrm{d}x_k) \mathrm{d}\mathbf{y}
\end{aligned}
$$

$$= \frac{p_n^{e(G)}}{a(G)} \int_{(\mathbb{R}^d)^{k-1}} \prod_{(\imath,\jmath)\in E(G)} H(x_\imath, x_\jmath) \mu^{\otimes(k-1)}(\mathrm{d}x_1, \ldots, \mathrm{d}x_{\ell-1}, \mathrm{d}x_{\ell+1}, \ldots, \mathrm{d}x_k), \tag{6.1}$$

where $\mathbf{x} = (x_1, \ldots, x_{\ell-1}, x, x_{\ell+1}, \ldots, x_k)$.

Proposition 6.1 *Suppose that the function f in (5.2) satisfies Assumption 4.1. Then, we have the following variance lower bound for subgraph counts:*

$$\mathrm{Var}[N_G] \geq \frac{C}{a(G)^2} \left(\frac{n!}{(n-2k+1)!} p_n^{2e(G)} + \frac{n!}{(n-k)!} p_n^{e(G)} \right), \tag{6.2}$$

where $C > 0$ is a constant independent of $n \geq 2k-1$.

Proof. We reuse the notation

$$I_\beta := \prod_{(\imath,\jmath)\in E_G} \mathbf{1}_{\{X_{\beta(\imath)} \sim X_{\beta(\jmath)}\}}, \quad \beta \in [n]^k_{\neq}.$$

Because the indicator random variable I_β, $\beta \in [n]^k_{\neq}$, indicates that there is a copy of G on $\{X_{\beta(1)}, \ldots, X_{\beta(k)}\}$, it is easy to see that for any $\alpha_1, \alpha_2 \in [n]^k_{\neq}$ we have

$$\mathbb{E}[I_{\alpha_1} \mid I_{\alpha_2} = 1] \geq \mathbb{E}[I_{\alpha_1}],$$

hence

$$\mathrm{Cov}(I_{\alpha_1}, I_{\alpha_2}) \geq 0, \quad \alpha_1, \alpha_2 \in [n]^k_{\neq}.$$

This further implies $\kappa(\mathbf{I}_\rho) \geq 0$ for any $\rho \in \mathrm{CNF}(2,k)$. For any $\rho \in \mathrm{CNF}(2,k)$ with $|\rho| = 2k-1$, there is only one block containing exactly two elements in ρ and the rest $2k-2$ blocks containing exactly one element. Without loss of generality, we assume $\rho \in \mathrm{CNF}(2,k)$ with $|\rho| = 2k-1$ and $\{(1,1),(2,t)\} \in \rho$ for some $t \in [k]$. Then, we have

$$\begin{aligned} \kappa(\mathbf{I}_\rho) &= \mathbb{E}\left[\left(\prod_{(\imath,\jmath)\in E_G} \mathbf{1}_{\{X_\imath \sim X_\jmath\}} \right) \left(\prod_{(\imath,\jmath)\in E_G} \mathbf{1}_{\{X'_\imath \sim X'_\jmath\}} \right) \right] - \left(\mathbb{E}\left[\prod_{(\imath,\jmath)\in E_G} \mathbf{1}_{\{X_\imath \sim X_\jmath\}} \right] \right)^2 \\ &= a(G)^2 \big(\mathbb{E}[f_{(1)}(X_1) f_{(t)}(X_1)] - \mathbb{E}[f_{(1)}(X_1)]^2 \big), \end{aligned}$$

where $\{X'_1, \ldots, X'_k\}$ are i.i.d. copies of $X \sim \mu$, independent of $\{X_1, \ldots, X_k\}$, with X'_t being replaced by X_1. From Assumption 4.1, we have

$$\sum_{\substack{\rho\in\mathrm{CNF}(2,k)\\ |\rho|=2k-1}} \kappa(\mathbf{I}_\rho) = a(G)^2 \left(\mathbb{E}\left[\left(\sum_{i=1}^k f_{(i)}(X_1) \right)^2 \right] - k^2 \mathbb{E}[f_{(1)}(X_1)]^2 \right)$$

$$= \quad a(G)^2 \mathrm{Var}\left[\sum_{i=1}^k f_{(i)}(X_1)\right] > 0.$$

On the other hand, combining the above with (6.1), we find that

$$\begin{aligned}
&\sum_{\substack{\rho\in\mathrm{CNF}(2,k)\\|\rho|=2k-1}} \kappa(\mathbf{I}_\rho) = a(G)^2 \mathrm{Var}\left[\sum_{i=1}^k f_{(i)}(X_1)\right] \\
&\quad = p_n^{2e(G)} \mathrm{Var}\left[\sum_{i=1}^k f_{(i)} \int_{(\mathbb{R}^d)^{k-1}} \prod_{(\imath,\jmath)\in E(G)} H(x_\imath, x_\jmath)\mu^{\otimes(k-1)}(\mathrm{d}x_1,\dots,\mathrm{d}x_{i-1},\mathrm{d}x_{i+1},\dots,\mathrm{d}x_k)\right] \\
&\quad = C_1 p_n^{2e(G)}, \qquad (6.3)
\end{aligned}$$

where

$$\begin{aligned}
C_1 &:= C_1(\mu, H) \\
&= \mathrm{Var}\left[\sum_{i=1}^k f_{(i)} \int_{(\mathbb{R}^d)^{k-1}} \prod_{(\imath,\jmath)\in E(G)} H(x_\imath, x_\jmath)\mu^{\otimes(k-1)}(\mathrm{d}x_1,\dots,\mathrm{d}x_{i-1},\mathrm{d}x_{i+1},\dots,\mathrm{d}x_k)\right] > 0
\end{aligned}$$

is a constant independent of $n \geq 1$ and $\mathbf{x} = (x_1, \dots, x_{\ell-1}, X_1, x_{\ell+1}, \dots, x_k)$.

Taking $j = 2$ in (5.6), together with (6.3), we obtain the following lower bound for the variance of subgraph counts:

$$\begin{aligned}
\mathrm{Var}[N_G] &= \frac{1}{a(G)^2} \sum_{r=k}^{2k-1} \frac{n!}{(n-r)!} \sum_{\substack{\rho\in\mathrm{CNF}(2,k)\\|\rho|=r}} \kappa(\mathbf{I}_\rho) \\
&\geq \frac{1}{a(G)^2}\left(\frac{n!}{(n-k)!} \sum_{\substack{\rho\in\mathrm{CNF}(2,k)\\|\rho|=k}} \kappa(\mathbf{I}_\rho) + \frac{n!}{(n-2k+1)!} \sum_{\substack{\rho\in\mathrm{CNF}(2,k)\\|\rho|=2k-1}} \kappa(\mathbf{I}_\rho)\right) \\
&\geq \frac{1}{a(G)^2}\left(\frac{n!}{(n-k)!} \kappa(\mathbf{I}_{\tilde\rho}) + \frac{n!}{(n-2k+1)!} \sum_{\substack{\rho\in\mathrm{CNF}(2,k)\\|\rho|=2k-1}} \kappa(\mathbf{I}_\rho)\right),
\end{aligned}$$

where $\tilde\rho$ is the partition $\tilde\rho = \big\{\{(1,1),(2,1)\},\dots,\{(1,k),(2,k)\}\big\}$ of $[2]\times[k]$ Next, for any $\beta = [\beta_1, \beta_2]^\top \in ([n]^k_{\neq})^2$ such that $\sqcap(\beta) = \tilde\rho$, we have

$$\kappa(\mathbf{I}_{\tilde\rho}) = \mathrm{Var}[I_\beta] = \theta p_n^r(1-\theta p_n^r) = \theta p_n^{e(G)}\left(1-\theta p_n^{e(G)}\right)$$

where

$$\theta := \int_{(\mathbb{R}^d)^k} \prod_{(\imath,\jmath)\in E(G)} H(x_\imath, x_\jmath)\mu^{\otimes k}(\mathrm{d}x_1,\dots,\mathrm{d}x_k),$$

hence

$$\begin{aligned}\operatorname{Var}[N_G] &\geq \frac{1}{a(G)^2}\left(\frac{n!}{(n-k)!}\theta p_n^{e(G)}\left(1-\theta p_n^{e(G)}\right)+\frac{n!}{(n-2k+1)!}C_1 p_n^{2e(G)}\right)\\ &\geq \frac{C_2}{a(G)^2}\left(\frac{n!}{(n-2k+1)!}p_n^{2e(G)}+\frac{n!}{(n-k)!}p_n^{e(G)}\right),\end{aligned}$$

with $C_2 := \min\left(C_1, \theta\left(1-\theta p_n^{e(G)}\right)\right)$. □

7 Growth rates of subgraph count cumulants

The general cumulant upper bound of the subgraph count N_G established in Theorem 5.2 does not yield an explicit asymptotic growth order. In this section, we perform a more detailed analysis of cumulant growth rates by identifying the leading terms of the form $n^r p_n^{d(j,k,r)}$ in (5.4), $r = k, \ldots, 1+(k-1)j$, where $d(j,k,r)$ is defined in (5.5), with in particular

$$d_k = e(G), \quad d_{1+(k-1)j} = je(G). \tag{7.1}$$

To this end, we start by deriving cumulant growth rates via the convex analysis of planar diagrams used in [LP24b] in the random-connection model and introduced in [JŁR00] for the Erdős-Rényi model. Firstly, we adopt some notation from [LP24b].

Definition 7.1 *Let G be a connected graph with $k \geq 2$ vertices. For $j \geq 1$, we let*

i) $\Sigma_j(G) := \{(x(\rho_G), y(\rho_G)) := (jk - v(\rho_G), je(G) - e(\rho_G)) : \rho \in \mathrm{CNF}(j,k)\}$, where for each $\rho \in \Pi([j] \times [k])$, ρ_G is the graph associated to ρ by Definition 2.3;

ii) and we let $\widehat{\Sigma}_j(G)$ denote the upper boundary of the convex hull of $\Sigma_j(G)$.

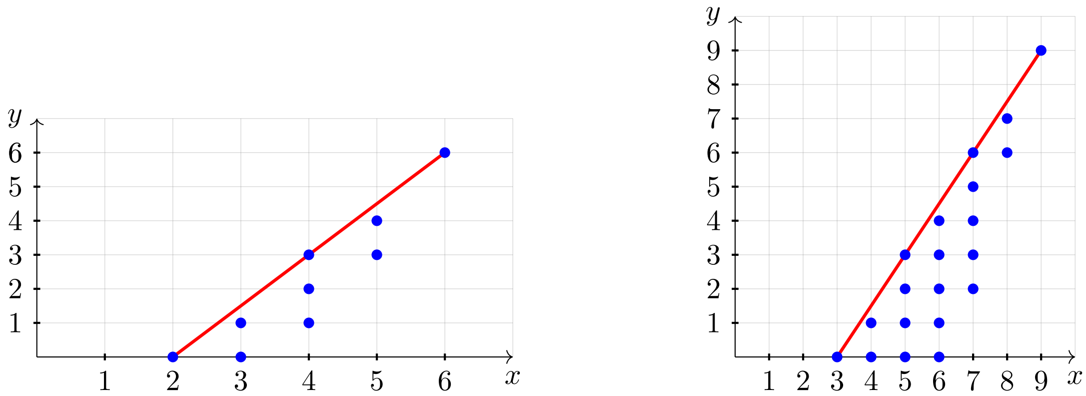
Remark 7.2 *According to (7.1), we can see that two endpoints of the upper boundary of the convex hull of $\Sigma_j(G)$ have coordinates $(j-1, 0)$ and $((j-1)k, (j-1)e(G))$.*

The following example was considered in [LP24b, Example 5.2] in the Poisson random-connection model.

Example 7.3 *Let $G = C_3$ be a triangle, i.e. $v(G) = e(G) = 3$. We have*

$$\begin{cases}\Sigma_2(C_3) = \{(3,3),(2,1),(1,0)\},\\ \Sigma_3(C_3) = \{(6,6),(5,4),(4,3),(5,3),(4,2),(4,1),(3,1),(3,0),(2,0)\},\\ \Sigma_4(C_3) = \{(9,9),(8,7),(7,6),(8,6),(7,5),(7,4),(6,4),(6,3),(5,3),(7,3),\\ \qquad (6,2),(5,2),(7,2),(6,1),(5,1),(4,1),(6,0),(5,0),(4,0),(3,0)\},\end{cases}$$

see Figure 5-(a) with $j=3$, $k=3$, $e(G)=3$, and Figure 5-(b) $j=4$, $k=3$, $e(G)=3$.



(a) Plot of $\Sigma_3(C_3)$. (b) Plot of $\Sigma_4(C_3)$.

Figure 5: Set $\Sigma_n(C_3)$ and upper boundary of its convex hull (in red) for $n=3,4$.

From Figure 5 we can also check that

$$\begin{aligned} \widehat{\Sigma}_3(C_3)\cap\Sigma_3(C_3) = &\quad \{(6,6),(4,3),(2,0)\}, \\ \widehat{\Sigma}_4(C_3)\cap\Sigma_4(C_3) = &\quad \{(9,9),(7,6),(5,3),(3,0)\}. \end{aligned}$$

We recall the following definition.

Definition 7.4 *[ŁR92], [JŁR00, pages 64-65] A graph G is strongly balanced if*

$$\frac{e(H)}{v(H)-1}\leq\frac{e(G)}{v(G)-1}, \qquad H\subset G. \tag{7.2}$$

Theorem 7.5 *Let G be a strongly balanced connected graph.*

a) If $n^{-(v(G)-1)/e(G)}\ll p_n$, then for n large enough we have

$$|\kappa_j(N_G)|\leq j^{j-1}j!^{v(G)}\frac{v(G)!^{j-1}}{a(G)^j}n^{1+(v(G)-1)j}p_n^{je(G)}, \quad j\geq 2. \tag{7.3}$$

b) If $p_n\ll n^{-(v(G)-1)/e(G)}$, then for n large enough we have

$$|\kappa_j(N_G)|\leq j^{j-1}j!^{v(G)}\frac{v(G)!^{j-1}}{a(G)^j}n^{v(G)}p_n^{e(G)}, \quad j\geq 2. \tag{7.4}$$

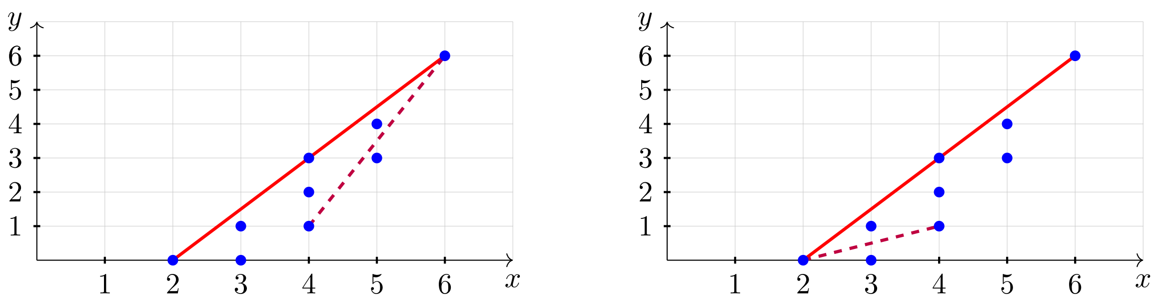
Proof. We let $k:=v(G)$. According to [LP24b, Proposition 6.4], if G is a strongly balanced graph, then the upper boundary of the convex hull of $\Sigma_j(G)$, $j\geq 2$, is a line segment, whose endpoints are $(j-1,0)$ and $((j-1)k,(j-1)e(G))$, see Remark 7.2 and Figure 6 for $j=k=3$. Then, for any $r\in\{k+1,\ldots,(k-1)j\}$, the asymptotic order of summands appearing in the upper bound of (5.4) is corresponding to a point $(jk-r, je(G)-d(j,k,r))$ denoted by the blue dots in Figure 6. For every such point $(jk - r, je(G) - d(j,k,r))$ located within the convex hull, comparing the ratio of a line segment that connects itself with the right endpoint $((j-1)k, (j-1)e(G))$ of the upper boundary of the convex hull with the ratio of the upper boundary, we have

$$\frac{(j-1)e(G) - (je(G) - d(j,k,r))}{(j-1)k - (jk - r)} \geq \frac{(j-1)e(G)}{(j-1)k - (j-1)} = \frac{e(G)}{k-1},$$

hence, choosing $j \geq 2$,

$$\frac{d(j,k,r) - e(G)}{r - k} \geq \frac{e(G)}{k-1}, \tag{7.5}$$

see for example the dashed line in Figure 6-(a) with $j = k = e(G) = 3$, $r = 5$ and $d(3,3,5) = 8$, in the framework of Example 2.4.



(a) Plot of $\Sigma_3(C_3)$. (b) Plot of $\Sigma_3(C_3)$.

Figure 6: Set $\Sigma_3(C_3)$ and upper boundary of its convex hull (in red).

Similarly, for any $r \in \{k+1, \ldots, (k-1)j\}$, when comparing ratio of line segment between point $(jk - r, je(G) - d(j,k,r))$ and the left endpoint of the upper boundary $(j-1, 0)$ with the ratio of the upper boundary $\widehat{\Sigma}_j(G)$, if $(jk - r, je(G) - d(j,k,r)) \in \widehat{\Sigma}_j(G)$, then we have

$$\frac{je(G) - d(j,k,r)}{jk - r - (j-1)} = \frac{(j-1)e(G)}{(j-1)k - (j-1)},$$

hence

$$\frac{je(G) - d(j,k,r)}{jk - r - (j-1)} = \frac{e(G)}{k-1}. \tag{7.6}$$

On the other hand, if $(jk - r, je(G) - d(j,k,r)) \notin \widehat{\Sigma}_j(G)$, we have

$$\frac{je(G) - d(j,k,r)}{jk - r - (j-1)} < \frac{e(G)}{k-1}, \tag{7.7}$$

see for example the dashed line in Figure 6-(b) with $j = k = e(G) = 3$, $r = 5$ and $d(3,3,5) = 8$, as in Example 2.4.

a) When $n^{-(k-1)/e(G)} \ll p_n$, using (7.6) and (7.7), for any $r \in \{k, \ldots, (k-1)j\}$ we have

$$\begin{aligned}
\frac{n^{1+(k-1)j} p_n^{je(G)}}{n^r p_n^{d(j,k,r)}} &= n^{1+(k-1)j-r} p_n^{je(G)-d(j,k,r)} \\
&= \left(n p_n^{\frac{je(G)-d(j,k,r)}{1+(k-1)j-r}} \right)^{1+(k-1)j-r} \\
&\gtrsim \left(n p_n^{\frac{e(G)}{k-1}} \right)^{1+(k-1)j-r} \\
&\gg 1. \qquad (7.8)
\end{aligned}$$

Therefore, after combining with (5.4), for n large enough we have

$$\begin{aligned}
|\kappa_j(N_G)| &\leq \frac{j^{j-1}}{a(G)^j} \sum_{r=k}^{1+(k-1)j} |\mathcal{C}(j,k,r)| n^r p_n^{d(j,k,r)} \\
&\leq \frac{j^{j-1}}{a(G)^j} n^{1+(k-1)j} p_n^{je(G)} \sum_{r=k}^{1+(k-1)j} |\mathcal{C}(j,k,r)| \\
&= \frac{j^{j-1}}{a(G)^j} n^{1+(k-1)j} p_n^{je(G)} |\mathrm{CNF}(j,k)| \\
&\leq j!^k k!^{j-1} \frac{j^{j-1}}{a(G)^j} n^{1+(k-1)j} p_n^{je(G)},
\end{aligned}$$

where the last inequality is from Lemma 2.5.

b) When $p_n \ll n^{-(k-1)/e(G)}$, for any $r \in \{k+1, \ldots, (k-1)j+1\}$, by (7.5) we have

$$\begin{aligned}
\frac{n^k p_n^{e(G)}}{n^r p_n^{d(j,k,r)}} &= \left(n p_n^{\frac{d(j,k,r)-e(G)}{r-k}} \right)^{-(r-k)} \\
&\gtrsim \left(n p_n^{\frac{e(G)}{k-1}} \right)^{-(r-k)} \\
&\gg 1. \qquad (7.9)
\end{aligned}$$

Hence, from (5.4), for n large enough we have

$$\begin{aligned}
|\kappa_j(N_G)| &\leq \frac{j^{j-1}}{a(G)^j} \sum_{r=k}^{1+(k-1)j} |\mathcal{C}(j,k,r)| n^r p_n^{d(j,k,r)} \\
&\leq \frac{j^{j-1}}{a(G)^j} n^k p_n^{e(G)} \sum_{r=k}^{1+(k-1)j} |\mathcal{C}(j,k,r)| \\
&= \frac{j^{j-1}}{a(G)^j} n^k p_n^{e(G)} |\mathrm{CNF}(j,k)| \\
&\leq j!^k k!^{j-1} \frac{j^{j-1}}{a(G)^j} n^k p_n^{e(G)}.
\end{aligned}$$

The proof is complete. $\square$

By inspection of the proof of Theorem 7.5, see (7.8) and (7.9) therein, we note that its conclusions also hold under the following alternative conditions.

Proposition 7.6 *Let G be a strongly balanced connected graph.*

a) If $n^{-(v(G)-1)/e(G)} \le p_n$, $n \ge 1$, then we have

$$|\kappa_j(N_G)| \le j^{j-1} j!^{v(G)} \frac{v(G)!^{j-1}}{a(G)^j} n^{1+(v(G)-1)j} p_n^{je(G)}, \quad j \ge 2, \ n \ge 1.$$

b) If $p_n \le n^{-(v(G)-1)/e(G)}$, $n \ge 1$, then we have

$$|\kappa_j(N_G)| \le j^{j-1} j!^{v(G)} \frac{v(G)!^{j-1}}{a(G)^j} n^{v(G)} p_n^{e(G)}, \quad j \ge 2, \ n \ge 1.$$

Proposition 7.7 is a straightforward consequence of Proposition 6.1. Note that (6.2) in Assumption 4.1 is valid in the binomial RCM, but it is not satisfied in the Erdős-Rényi model. Indeed, the principal degree of f equals 2 in the Erdős-Rényi model, see § 9 in [JN91].

Proposition 7.7 *Let G be a strongly balanced connected graph. Suppose that the function f in (5.2) satisfies Assumption 4.1.*

a) If $n^{-(v(G)-1)/e(G)} \ll p_n$, we have the lower bound

$$\kappa_2(N_G) \ge \frac{C}{a(G)^2} \frac{n!}{(n+1-2v(G))!} p_n^{2e(G)}, \quad j \ge 1, \tag{7.10}$$

where $C > 0$ is a constant independent of $n \ge 2v(G) - 1$.

b) If $p_n \ll n^{-(v(G)-1)/e(G)}$, we have the lower bound

$$\kappa_2(N_G) \ge \frac{C}{a(G)^2} \frac{n!}{(n-v(G))!} p_n^{e(G)}, \quad j \ge 1, \tag{7.11}$$

where $C > 0$ is a constant independent of $n \ge 2v(G) - 1$.

As a consequence of Theorem 7.5 and 7.7, we have the following corollary in the binomial RCM.

Corollary 7.8 *Let G be a strongly balanced connected graph. Suppose that the function f in (5.2) satisfies Assumption 4.1. Then, for $n \geq 1$ large enough, the j-th cumulant of the normalized subgraph count*

$$\overline{N}_G := \frac{N_G - \kappa_1(N_G)}{\sqrt{\kappa_2(N_G)}}$$

satisfies

$$\kappa_j\left(\overline{N}_G\right) \leq \begin{cases} \dfrac{j!^{1+v(G)}}{(Cn)^{j/2-1}} & \text{when} \quad n^{-(v(G)-1)/e(G)} \ll p_n, \\ \dfrac{j!^{1+v(G)}}{\left(Cn^{v(G)}p_n^{e(G)}\right)^{j/2-1}} & \text{when} \quad p_n \ll n^{-(v(G)-1)/e(G)}, \end{cases}$$

$j \geq 3$, where C is a constant independent of $n \geq 1$.

Proof. We let $k = v(G)$. When $n^{-(k-1)/e(G)} \ll p_n$, from the inequalities (4.9), (4.10), (7.3) and (7.10), we have

$$\begin{aligned}
\kappa_j\left(\overline{N}_G\right) &= \frac{\kappa_j(N_G)}{\kappa_2(N_G)^{j/2}} \\
&\leq j!^k k!^{j-1} \frac{j^{j-1}}{a(G)^j} n^{1+(k-1)j} p_n^{je(G)} \left(\frac{C}{a(G)^2} \frac{n!}{(n-2k+1)!} p_n^{2e(G)}\right)^{-j/2} \\
&= j^{j-1} j!^k k!^{j-1} n^{1+(k-1)j} \left(C \frac{n!}{(n-2k+1)!}\right)^{-j/2} \\
&\leq j!^{k+1} e^j k!^{j-1} n^{1+(k-1)j} \left(C \left(\frac{n}{2}\right)^{2k-1}\right)^{-j/2} \\
&= \frac{j!^{k+1}}{n^{(j-2)/2}} e^j k!^{j-1} C^{-j/2} 2^{j(k-1/2)}.
\end{aligned}$$

Similarly, when $p_n \ll n^{-(k-1)/e(G)}$, from the inequality

$$\frac{n!}{(n-k)!} \geq (n-k+1)^k = \left(1 - \frac{k-1}{n}\right)^k n^k \geq \left(\frac{3n}{4}\right)^k$$

and (7.4)-(7.11), we have

$$\begin{aligned}
\kappa_j\left(\overline{N}_G\right) &= \frac{\kappa_j(N_G)}{\kappa_2(N_G)^{j/2}} \\
&\leq j!^k k!^{j-1} \frac{j^{j-1}}{a(G)^j} n^k p_n^{e(G)} \left(\frac{C}{a(G)^2} \frac{n!}{(n-k)!} p_n^{e(G)}\right)^{-j/2} \\
&= j^{j-1} j!^k k!^{j-1} n^k p_n^{e(G)} \left(C \frac{n!}{(n-k)!} p_n^{e(G)}\right)^{-j/2} \\
&\leq \frac{(j-1)!}{\sqrt{2\pi j}} e^j j!^k k!^{j-1} n^k p_n^{e(G)} \left(C \left(\frac{3n}{4}\right)^k p_n^{e(G)}\right)^{-j/2}
\end{aligned}$$

$$
\begin{aligned}
&\leq e^j j!^{k+1} k!^{j-1} n^k p_n^{e(G)} \left(C \left(\frac{3n}{4} \right)^k p_n^{e(G)} \right)^{-j/2} \\
&= \frac{j!^{k+1} e^j k!^{j-1}}{\left(n^k p_n^{e(G)}\right)^{(j-2)/2}} \left(C \left(\frac{3}{4} \right)^k \right)^{-j/2}.
\end{aligned}
$$

□

From Corollary 7.8, we can see that the cumulants of the normalized subgraph count $\overline{N}_G$ satisfy the Statulevičius growth condition (A.2) with $\gamma := v(G)$ and

$$
\Delta_n := \begin{cases} (Cn)^{1/2} & \textit{when} \quad n^{-(v(G)-1)/e(G)} \ll p_n, \\ \left(C n^{v(G)} p_n^{e(G)}\right)^{1/2} & \textit{when} \quad p_n \ll n^{-(v(G)-1)/e(G)}, \end{cases}
$$

where C is a constant independent of $n \geq 1$. Therefore, from Proposition A.1-i) we have the following results.

Corollary 7.9 *(Kolmogorov bound). Let G be a strongly balanced connected graph. Suppose that the function f in (5.2) satisfies Assumption 4.1. For $n \geq 1$ large enough the normalized subgraph count $\overline{N}_G$ satisfies*

$$
\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\overline{N}_G \leq x\right) - \Phi(x) \right| \leq \begin{cases} \dfrac{C}{n^{1/(2+4v(G))}} & \textit{when} \quad n^{-(v(G)-1)/e(G)} \ll p_n, \\ \dfrac{C}{\left(n^{v(G)} p_n^{e(G)}\right)^{1/(2+4v(G))}} & \textit{when} \quad p_n \ll n^{-(v(G)-1)/e(G)}. \end{cases}
$$

for some $C > 0$.

By Corollary 7.8 and Proposition A.1-ii), we also have the following result.

Corollary 7.10 *(Moderate deviation principle). Let G be a strongly balanced connected graph. Suppose that the function f in (5.2) satisfies Assumption 4.1. Let $(a_n)_{n\geq 1}$ be a sequence of real numbers tending to infinity, and such that*

$$
\begin{cases} a_n \ll n^{1/(2+4v(G))} & \textit{when} \quad n^{-(v(G)-1)/e(G)} \ll p_n, \\ a_n \ll \left(n^{v(G)} p_n^{e(G)}\right)^{1/(2+4v(G))} & \textit{when} \quad p_n \ll n^{-(v(G)-1)/e(G)}. \end{cases}
$$

Then, $\left(a_n^{-1}\overline{N}_G\right)_{n\geq 1}$ satisfies a moderate deviation principle with speed a_n^2 and rate function $x^2/2$.

Another direct consequence of Theorem 7.5 is the following threshold phenomenon of subgraph containment in the binomial RCM.

Corollary 7.11 *Let G be a strongly balanced connected graph, and suppose that the function f in (5.2) satisfies Assumption 4.1. We have*

a) $\lim_{n\to\infty} \mathbb{P}(N_G = 0) = 1$ *if* $p_n \ll n^{-v(G)/e(G)}$,

b) $\lim_{n\to\infty} \mathbb{P}(N_G = 0) = 0$ *if* $p_n \gg n^{-v(G)/e(G)}$.

Proof. We let $k = v(G)$. When $p_n \gg n^{-k/e(G)}$, from (7.3) we have

$$\kappa_2(N_G) \leq \frac{2^{k+1}k!}{a(G)^2} n^{1+2(k-1)} p_n^{2e(G)},$$

hence by (5.3) we have

$$\lim_{n\to\infty} \frac{\kappa_2(N_G)}{\kappa_1(N_G)^2} = 0,$$

and

$$\lim_{n\to\infty} \frac{(\mathbb{E}[N_G])^2}{\mathbb{E}[N_G^2]} = \lim_{n\to\infty} \frac{\kappa_1(N_G)^2}{\kappa_2(N_G) + \kappa_1(N_G)^2} = 1.$$

On the other hand, if $p_n \ll n^{-k/e(G)}$, from (5.3), we have

$$\lim_{n\to\infty} \mathbb{E}(N_G) = 0.$$

We conclude by the first and second moment methods, see [JŁR00, Page 54], which state that

$$\frac{(\mathbb{E}[N_G])^2}{\mathbb{E}[N_G^2]} \leq \mathbb{P}(N_G > 0) \leq \mathbb{E}[N_G]. \tag{7.12}$$

□

A Cumulant method

Given $(X_l)_{l\geq 1}$ a sequence of random variables, for any subset b of $\{1, \ldots, n\}$ we consider a family

$$\mathbf{X}_b = (X_l)_{l\in b}$$

indexed by $b \subset [n]$. Taking $b := \{j_1, \ldots, j_k\}$, the joint characteristic function of the vector $\mathbf{X}_b$ is defined as

$$\varphi_{\mathbf{X}_b}(t_1, \ldots, t_k) := \mathbb{E}\left[\exp\left(i \sum_{\ell=1}^{k} t_\ell X_{j_\ell}\right)\right],$$

and the the joint cumulant of $\mathbf{X}_b$ is defined as

$$\kappa(\mathbf{X}_b) = (-i)^k \frac{\partial^k}{\partial t_1 \cdots \partial t_k} \ln \varphi_{\mathbf{X}_b}(t_1, \ldots, t_k)|_{t_1=\cdots=t_k=0}.$$

For any random variable ξ, we let $\xi'_b := (\xi, \ldots, \xi)$ denote the random vector with $|b|$ entries identical to ξ, and write $\kappa_n(\xi) := \kappa(\xi'_{[n]})$ for the n-*th* order cumulant of ξ. From Theorem 1 in [Luk55], see also [LS59], [Mal80], we have the relation

$$\kappa(\mathbf{X}_b) = \sum_{\sigma \in \Pi(b)} (-1)^{s-1}(s-1)! \prod_{a \in \sigma} \mathbb{E}\left[\prod_{l \in a} X_l\right] \tag{A.1}$$

between the joint moments and cumulants of the random vector $\mathbf{X}_b$, $b \subset [n]$. The following results are summarized from the "main lemmas" in Chapter 2 of [SS91] and [DE13].

Proposition A.1 *Let $(X_n)_{n\geq 1}$ be a family of random variables with mean zero and unit variance for all $n \geq 1$. Suppose that for all $n \geq 1$, all moments of the random variable X_n exist and that the cumulants of X_n satisfy*

$$|\kappa_j(X_n)| \leq \frac{(j!)^{1+\gamma}}{(\Delta_n)^{j-2}}, \qquad j \geq 3, \tag{A.2}$$

where $\gamma \geq 0$ depends only on $n \geq 1$, while $\Delta_n \in (0, \infty)$ may depend on $n \geq 1$. Then, the following assertions hold.

i) (Kolmogorov bound, [SS91, Corollary 2.1] and [DJS22, Theorem 2.4]) One has

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(X_n \leq x) - \Phi(x)| \leq \frac{C(\gamma)}{(\Delta_n)^{1/(1+2\gamma)}},$$

for some $C(\gamma) > 0$ depending only on γ.

ii) (Moderate deviation principle, [DE13, Theorem 1.1] and [DJS22, Theorem 3.1]). Let $(a_n)_{n\geq 1}$ be a sequence of real numbers tending to infinity, and such that

$$a_n \ll (\Delta_n)^{1/(1+2\gamma)}.$$

Then, $(a_n^{-1} X_n)_{n\geq 1}$ satisfies a moderate deviation principle with speed a_n^2 and rate function $x^2/2$.

iii) (Concentration inequality, corollary of [SS91, Lemma 2.4] and [DJS22, Theorem 2.5]). For any $x \geq 0$ and sufficiently large $n \geq 1$,

$$\mathbb{P}(|X_n| \geq x) \leq 2\exp\left(-\frac{1}{4}\min\left(\frac{x^2}{2^{1+\gamma}}, (x\Delta_n)^{1/(1+\gamma)}\right)\right).$$

iv) (Normal approximation with Cramér corrections, [SS91, Lemma 2.3]). There exists a constant $c > 0$ such that for all $n \geq 1$ and $x \in (0, c(\Delta_n)^{1/(1+2\gamma)})$ we have

$$\begin{aligned} \frac{\mathbb{P}(X_n \geq x)}{1 - \Phi(x)} &= \left(1 + O\left(\frac{x+1}{(\Delta_n)^{1/(1+2\gamma)}}\right)\right) \exp\left(\widetilde{L}(x)\right), \\ \frac{\mathbb{P}(X_n \leq -x)}{\Phi(-x)} &= \left(1 + O\left(\frac{x+1}{(\Delta_n)^{1/(1+2\gamma)}}\right)\right) \exp\left(\widetilde{L}(-x)\right), \end{aligned}$$

where $\widetilde{L}(x)$ is related to the Cramér-Petrov series.

Acknowledgements

We thank the referees for useful suggestions. This research is supported by the Ministry of Education, Singapore, under its Tier 1 Grant RG103/23.

References

[AcdOG25] J.D. Alvarado, L. Gonçalves de Oliveira, and S. Griffiths. Moderate deviations of triangle counts in sparse Erdős-Rényi random graphs $G(n,m)$ and $G(n,p)$. *Probab. Theory Relat. Fields*, 2025.

[BCJ23] B. Bhattacharya, A. Chatterjee, and S. Janson. Fluctuations of subgraph counts in graphon based random graphs. *Combin. Probab. Comput.*, 32(3):428–464, 2023.

[BDMM24] B.B. Bhattacharya, S. Das, S. Mukherjee, and S. Mukherjee. Fluctuations of quadratic chaos. *Commun. Math. Phys.*, 405(237), 2024.

[CGR16] M. Coulson, R. E. Gaunt, and G. Reinert. Poisson approximation of subgraph counts in stochastic block models and a graphon model. *ESAIM Probab. Stat.*, 20:131–142, 2016.

[CT22] V. H. Can and K. D. Trinh. Random connection models in the thermodynamic regime: central limit theorems for add-one cost stabilizing functionals. *Electron. J. Probab.*, 27:1–40, 2022.

[DE09] H. Döring and P. Eichelsbacher. Moderate deviations in a random graph and for the spectrum of bernoulli random matrices. *Electron. J. Probab.*, 14:2636–2656, 2009.

[DE13] H. Döring and P. Eichelsbacher. Moderate deviations via cumulants. *J. Theoret. Probab.*, 26:360–385, 2013.

[DF14] L. Devroye and N. Fraiman. Connectivity of inhomogeneous random graphs. *Random Structures Algorithms*, 45(3):408–420, 2014.

[DJS22] H. Döring, S. Jansen, and K. Schubert. The method of cumulants for the normal approximation. *Probab. Surv.*, 19:185–270, 2022.

[DM83] E. B. Dynkin and A. Mandelbaum. Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Ann. Statist.*, 11(3):739–745, 1983.

[DST16] L. Decreusefond, M. Schulte, and C. Thäle. Functional Poisson approximation in Kantorovich-Rubinstein distance with applications to U-statistics and stochastic geometry. *Ann. Probab.*, 44(3):2147–2197, 2016.

[ER23] P. Eichelsbacher and B. Rednoß. Kolmogorov bounds for decomposable random variables and subgraph counting by the Stein-Tikhomirov method. *Bernoulli*, 29(3):1821–1848, 2023.

[FMN16] V. Féray, P.-L. Méliot, and A. Nikeghbali. *Mod-ϕ Convergence Normality Zones and Precise Deviations*. SpringerBriefs in Probability and Mathematical Statistics. Springer Cham, 2016.

[HHO25] N. Heerten, C. Hirsch, and M. Otto. Moderate deviations for weight-dependent random connection models. *J. Appl. Probab.*, pages 1–23, 2025.

[Hoe61] W. Hoeffding. The strong law of large numbers for U-statistics. Technical report, North Carolina State University. Dept. of Statistics, 1961.

[HPČ21] J. Hladký, C. Pelekis, and M. Čileikis. A limit theorem for small cliques in inhomogeneous random graphs. *J. Graph Theory*, 97:578–599, 2021.

[Jan97] S. Janson. *Gaussian Hilbert spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.

[JŁR00] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[JN91] S. Janson and K. Nowicki. The asymptotic distributions of generalized U-statistics with applications to random graphs. *Probab. Theory Related Fields*, 90(3):341–375, 1991.

[Kho08] O. Khorunzhiy. On connected diagrams and cumulants of Erdős-Rényi matrix models. *Comm. Math. Phys.*, 282:209–238, 2008.

[KR21] G. Kaur and A. Röllin. Higher-order fluctuations in dense random graph models. *Electron. J. Probab.*, 25:1–36, 2021.

[LNS21] G. Last, F. Nestmann, and M. Schulte. The random connection model and functions of edge-marked Poisson processes: second order properties and normal approximation. *Ann. Appl. Probab.*, 31(1):128–168, 2021.

[LP24a] Q. Liu and N. Privault. Normal approximation of subgraph counts in the random-connection model. *Bernoulli*, 30:3224–3250, 2024.

[LP24b] Q. Liu and N. Privault. Normal to Poisson phase transition for subgraph counting in the random-connection model. Preprint arXiv:2409.16222, 37 pages, 2024.

[ŁR92] T. Łuczak and A. Ruciński. Convex hulls of dense balanced graphs. *J. Comput. Appl. Math.*, 41:205–213, 1992.

[LRP17] R. Lachièze-Rey and G. Peccati. New Berry-Esseen bounds for functionals of binomial point processes. *Ann. Appl. Probab.*, 27(4):1992–2031, 2017.

[LS59] V.P. Leonov and A.N. Shiryaev. On a method of calculation of semi-invariants. *Theory Probab. Appl.*, 4:319–329, 1959.

[Luk55] E. Lukacs. Applications of Faà di Bruno's formula in mathematical statistics. *Amer. Math. Monthly*, 62:340–348, 1955.

[LX25] Q. Liu and A. Xia. Poisson approximation for subgraph counts in general random-connection models. Preprint, 29 pages, 2025.

[Mal80] V.A. Malyshev. Cluster expansions in lattice models of statistical physics and the theory of quantum fields. *Russian Mathematical Surveys*, 35(1):1–62, 1980.

[MM91] V.A. Malyshev and R.A. Minlos. *Gibbs Random Fields*, volume 44 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1991.

[Pen03] M. Penrose. *Random Geometric Graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, 2003.

[Pen18] M. Penrose. Inhomogeneous random graphs, isolated vertices, and Poisson approximation. *J. Appl. Probab.*, 55(1):112–136, 2018.

[PS18] N. Privault and G. Serafin. Stein approximation for functionals of independent random sequences. *Electron. J. Probab.*, 23:Paper No. 4, 1–34, 2018.

[PT11] G. Peccati and M. Taqqu. *Wiener Chaos: Moments, Cumulants and Diagrams: A survey with Computer Implementation.* Bocconi & Springer Series. Springer, 2011.

[SS91] L. Saulis and V.A. Statulevičius. *Limit Theorems for Large Deviations*, volume 73 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1991.

[ST24] M. Schulte and C. Thäle. Moderate deviations on Poisson chaos. *Electron. J. Probab.*, 29:no. 146, 1–27, 2024.

[Zha22] Z.S. Zhang. Berry-Esseen bounds for generalized U-statistics. *Electron. J. Probab.*, 27:1–36, 2022.