Electronic Journal of Probability

Electron. J. Probab. **30** (2025), article no. 170, 1–32.
ISSN: 1083-6489 https://doi.org/10.1214/25-EJP1434

Normal to Poisson phase transition for subgraph counting in the random-connection model*

Qingwei Liu† Nicolas Privault‡

Abstract

We consider the limiting behavior of the count of subgraphs isomorphic to a graph G with $m \geq 0$ fixed endpoints (or roots) in the random-connection model, as the intensity λ of the underlying Poisson point process tends to infinity. When connection probabilities are of order $\lambda^{-\alpha}$ we identify a phase transition phenomenon depending on a critical decay rate $\alpha_m^*(G) > 0$ such that normal approximation for subgraph counts holds when $\alpha \in (0, \alpha_m^*(G))$, and a Poisson limit result holds if $\alpha = \alpha_m^*(G)$. Our approach relies on cumulant growth rates derived by the convex analysis of planar diagrams that enumerate the partitions involved in cumulant identities. As a result, by the cumulant method we obtain normal approximation results with convergence rates in the Kolmogorov distance, and a Poisson limit theorem, for subgraph counts.

Keywords: random-connection model; Poisson point process; random graphs; subgraph counting; rooted graphs; cumulant method; phase transition.
MSC2020 subject classifications: 60F05; 60D05; 05C80; 60G55.
Submitted to EJP on November 9, 2024, final version accepted on October 29, 2025.

1 Introduction

The first instance of a threshold phenomenon in random graphs was observed in the Erdős-Rényi model for the containment of balanced graphs, see [ER61]. In [Bol81b, Theorem 1], a Poisson limit theorem for the counting of strictly balanced subgraphs was proved at the threshold. Poisson approximation under the total variation distance has been proved in [Bar82] via the Stein-Chen method, see also [BHJ92, Chapter 5]. Phase transition phenomena for inhomogeneous random graphs have been studied for the growth rate of the giant component in [BJR07], and for connectivity thresholds in [DF14].

Consider an Erdős-Rényi random graph on n vertices, with independent connection probability p_n such that $p_n = c/n^\alpha$ for some $c > 0$ and $\alpha > 0$. Necessary and sufficient conditions for the asymptotic normality of the (normalized) count $\overline{N}_G$ of subgraphs isomorphic to a fixed graph G were obtained in [Ruc88], as follows.

*This research is supported by the Ministry of Education, Singapore, under its Tier 1 Grant RG103/23.
†University of Melbourne, Australia. E-mail: qingwei.liu@unimelb.edu.au
‡Nanyang Technological University, Singapore. E-mail: nprivault@ntu.edu.sg

- When
$$0 < \alpha < \min_{H \subset G} \frac{v(H)}{e(H)}, \tag{1.1}$$
the normalized subgraph count $\overline{N}_G$ converges to a normal random variable, where $v(H)$ and $e(H)$ respectively denote the counts of vertices and edges of any subgraph H of G.
- When
$$\alpha = \min_{H \subset G} \frac{v(H)}{e(H)},$$
the rescaled subgraph count converges to a Poisson distribution provided that the graph G is strictly balanced, see Definition 2.1-(1), and
$$\lim_{n\to\infty} n p_n^{1/\alpha} = c > 0, \tag{1.2}$$
see [Bol81a], [KR83], and Theorem 3.19 in [JŁR00].

More recently, explicit convergence rates in the Wasserstein distance for subgraph counts have been obtained in [BKR89], see also [PS20] and references therein for rates in the Kolmogorov distance using the discrete Malliavin calculus.

Poisson and compound Poisson approximation of subgraph counts were obtained together with convergence rates in [CGR16] and [CGR18] for the stochastic block model, which can be viewed as a special case of the graphon-based random-connection model. Normal approximation results for subgraph counts in graphon-based random graphs were obtained in [KR21], [Zha22], [BCJ23], and [LP25a] via the determination of quantitative bounds and higher order fluctuations.

This paper considers the counting of subgraphs isomorphic to a fixed graph in the Poisson random-connection model (RCM) $G_\varphi(\eta)$, which is a random graph whose vertex set is given by a Poisson point process η with intensity Λ on $\mathbb{R}^d$, $d \geq 1$, and where every pair of vertices is randomly connected with a location-dependent probability given by a connection function $\varphi : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$. The random-connection model can be regarded as a unified framework containing several classical random graph models as particular cases.

In particular, when φ is a function of the distance between pairs of points of η, i.e. $\varphi(x,y) := \phi(\|x-y\|)$ for some measurable function $\phi : \mathbb{R}_+ \to [0,1]$, the resulting graph is also known as a soft random geometric graph, see [Pen91], [Pen16], and [LNS21]. When ϕ takes the form $\phi(u) = \mathbf{1}_{\{u \leq \varepsilon\}}$, for some $\varepsilon > 0$, the random-connection model becomes a random geometric graph, c.f. the monograph [Pen03], in which a pair of vertices is connected by an edge if and only if the distance between them is less than the fixed threshold ε. For another example, when the underlying point process η is a binomial point process and the connection function $\varphi(x,y) \equiv p$ is constant for some $p \in (0,1)$, the resulting graph is the Erdős-Rényi random graph, c.f. [ER59] and [Gil59]. When the connection function $\varphi(x,y)$ is symmetric on $[0,1] \times [0,1]$ it is also called a graphon, and the corresponding random-connection model arises as a limit of dense graph sequences, see [LS06], [Lov12], [Zha22], [BCJ23].

In the Poisson random-connection model, phase transition phenomena have been observed in [MR96], where a critical point Poisson intensity parameter has been identified for the occurrence of percolation. Poisson approximation results for edge counts and for subgraphs of the same order were established in [Pen18], and normal approximation result have been obtained for subgraph counts in [CT22]. In [LP24], normal approximation for the subgraph counts in the Poisson random-connection model as the intensity of the underlying Poisson point process goes to infinity has been established, together with convergence rates under the Kolmogorov distance, through combinatorial arguments and the cumulant method, see also [GT18], [Jan19], and [DE13], [ST24], [HHO25] for moderate deviations. By expressing the cumulants of subgraph counts as sums over non-flat and connected partition diagrams, cumulant growth has been analyzed under different limiting regimes, leading to the asymptotic normality of connected subgraph counts in the dilute case, and for trees in the sparse case.

In this paper, we derive convergence rates in the Kolmogorov distance for the normal approximation of subgraph counts in the Poisson random-connection model, and we identify a critical threshold at which Poisson convergence occurs. To the best of our knowledge, this is the first time that such a phase transition, from normal to Poisson limit theorems for subgraph counts, has been observed in the Poisson random-connection model. In addition, we include the counting of subgraphs containing one or more endpoints, defined as roots placed as arbitrary deterministic locations. This extends previous approaches to the counting of rooted subgraphs in the Erdős-Rényi and inhomogeneous random graph models, see [RV86] and [Mau24]. In applications to e.g. wireless networks, endpoints can model physical devices placed at given fixed locations, such as roadside units in vehicular networks, see, e.g., [NZZ+11], [ZCY+12], and [KGKLN23].

We investigate the asymptotic behavior of subgraph counts in a Poisson random-connection model. By allowing the connection probability to vary as the intensity of the underlying Poisson point process tends to infinity, we identify a phase transition where the limiting distribution shifts between normal and Poisson, whenever the graph G satisfies a balance condition (2.6). Similar to the Erdős-Rényi graph, while this balance condition (2.6) may not be necessary for asymptotic normality, we believe it is essential for the emergence of a Poisson limit, see Remark 6.4.

The counting of subgraphs containing one or more fixed endpoints, which covers the counting of rooted subgraphs [RV86], [Mau24] in the case of a single endpoint, has been considered in [LP25b] in the random-connection model. However, the analyses of cumulant growths in [LP24] and [LP25b], were constrained by the use of partitions of maximal cardinality, which resulted in an incomplete characterization of the limiting regimes that ensure asymptotic normality. In the present paper, we overcome those restrictions via a detailed analysis of the behavior of the count N_G of subgraphs isomorphic to a given connected graph G in the random-connection model with and without fixed endpoints.

In comparison with [LP24] and [LP25b], the present paper provides a unified analysis of normal approximation for subgraph counts in terms of a single threshold and in the presence of endpoints, without restriction to specific dilute and sparse regimes, see Theorem 3.3 and Corollary 3.4. This analysis is performed in terms of subgraph densities, and it covers the Poisson convergence regime which is shown to hold at the parameter threshold, see Theorem 3.6.

For this, we develop new combinatorial tools in the random-connection model, based on the convex analysis of subgraph plots introduced for the Erdős-Rényi model in [JŁR00], which we use to analyse the partitions involved in the representation of subgraph count cumulants. This yields an exhaustive analysis of asymptotic normalized cumulant growth on a random-connection model $G_\varphi(\eta \cup \{y_1, \dots, y_m\})$ including endpoints $y_1, \dots, y_m$, with intensity measure of the form $\lambda \cdot \mu$, $\lambda > 0$, in Proposition 7.1, where μ is a diffuse sigma-finite measure on $\mathbb{R}^d$ and the connection function $\varphi : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ is rescaled as $\varphi_\lambda := c_\lambda \cdot \varphi$, $\lambda > 0$.

Growth rates for the cumulants of normalized subgraph counts are then obtained on the random-connection model $G_\varphi(\eta \cup \{y_1, \dots, y_m\})$ in Theorem 3.3 under Assumptions 3.2-(i)-(iii) and the balance condition (2.6). Given two functions $f(\lambda)$ and $g(\lambda) > 0$, we write

- $f(\lambda) = O(g(\lambda))$, or $f(\lambda) \lesssim g(\lambda)$, if $\limsup_{\lambda\to\infty} f(\lambda)/g(\lambda) < \infty$,

- $f(\lambda) = \Omega(g(\lambda))$, or $f(\lambda) \gtrsim g(\lambda)$, if $\liminf_{\lambda\to\infty} f(\lambda)/g(\lambda) > 0$,
- $f(\lambda) \asymp g(\lambda)$ if $f(\lambda) = O(g(\lambda))$ and $f(\lambda) = \Omega(g(\lambda))$,
- $f(\lambda) \sim g(\lambda)$ if $\lim_{\lambda\to\infty} f(\lambda)/g(\lambda) = 1$,
- $f(x) \ll g(x)$, or $g(x) \gg f(x)$, if $f(x) \geq 0$ and $f(x)/g(x) \to 0$.

with the convention $0/0 = 0$. For G a graph with $v(G) = r + m$ vertices including m endpoints, we have the following consequences of Theorem 3.3.

- We show in Corollary 3.4 that when
$$1 \gtrsim c_\lambda \gg \lambda^{-\min(r/e(G),1/a_m(G))},$$
the normalized subgraph count $\overline{N}_G$ converges to a normal random variable as λ tends to infinity, where $a_m(G)$ is defined in (2.4) and depends on endpoint connectivity. As a consequence, when c_λ takes the form $c_\lambda \asymp \lambda^{-\alpha}$ as λ tends to infinity for some $\alpha > 0$, we extend the thresholds (1.1)-(1.2) in [JŁR00] from the Erdős-Rényi model to the random-connection model, by showing that, under the balance condition (2.6), normal approximation holds for the normalized subgraph count $\overline{N}_G$ provided that
$$0 < \alpha < \alpha_m^*(G) := \min\left(\frac{r}{e(G)}, \frac{1}{a_m(G)}\right),$$
i.e.
$$0 < \alpha < \alpha_m^*(G) = \frac{r}{e(G)} \tag{1.3}$$
when $m = 0$ or $m = 1$.
- In Theorem 3.5, we derive convergence rates under the Kolmogorov distance, together with a moderate deviation principle, concentration inequalities and a normal approximation result with Cramér correction. In particular, when $c_\lambda \asymp \lambda^{-\alpha}$, we obtain the Kolmogorov bounds
$$\sup_{x\in\mathbb{R}} |\mathbb{P}_\lambda(\overline{N}_G \leq x) - \Phi(x)| \leq \begin{cases} \dfrac{C}{\lambda^{(1-\alpha a_m(G))/(4r-2)}} & \text{if } 0 < \alpha \leq \dfrac{r-1}{e(G) - a_m(G)}, \\[2ex] \dfrac{C}{\lambda^{(r-\alpha e(G))/(4r-2)}} & \text{if } \dfrac{r-1}{e(G) - a_m(G)} \leq \alpha < \alpha_m^*(G), \end{cases}$$
where Φ is the cumulative distribution of the standard normal distribution and $C > 0$ is a constant depending only on $r \geq 2$. When G has no endpoints $(m = 0)$ and is strongly balanced, we have
$$\sup_{x\in\mathbb{R}} |\mathbb{P}_\lambda(\overline{N}_G \leq x) - \Phi(x)| \leq \begin{cases} \dfrac{C}{\lambda^{1/(4v(G)-2)}} & \text{if } 0 < \alpha \leq \dfrac{v(G)-1}{e(G)}, \\[2ex] \dfrac{C}{\lambda^{(v(G)-\alpha e(G))/(4v(G)-2)}} & \text{if } \dfrac{v(G)-1}{e(G)} \leq \alpha < \alpha_0^*(G) = \dfrac{v(G)}{e(G)}, \end{cases}$$
which extends Corollary 7.1 of [LP24] beyond the dilute regime considered there, and also Corollary 7.2 therein without restriction to trees.

- Under the condition $a_m(G)r \leq e(G)$, Poisson convergence holds for N_G by Theorem 3.6 in the boundary case

$$\alpha = \alpha_m^*(G) = \frac{r}{e(G)},$$

- Finally, by Theorem 3.7-(a), N_G converges to zero in probability if $a_m(G)r \leq e(G)$ and

$$\alpha > \alpha_m^*(G) = \frac{r}{e(G)}.$$

Remark 1.1. a) In the case of rooted subgraph counting, i.e. when $m = 1$ with a single endpoint, Condition (1.3) is consistent with the Property (P) page 261 of [RV86] in the Erdős-Rényi model, and with the asymptotic normality condition in Theorem 1 of [Mau24] in inhomogeneous random graphs.

b) In the absence of endpoints $(m = 0)$, Condition (2.6) means that G should be strongly balanced, and (1.3) reads

$$0 < \alpha < \alpha_0^*(G) = \min_{H \subset G} \frac{v(H)}{e(H)} = \frac{v(G)}{e(G)},$$

which coincides with (1.1), see Definition 2.1-(1).

c) We note that in the random-connection model, our results require a strong balance condition of the form (2.6), which is not needed in the Erdős-Rényi model, see [Ruc88], [Bol81a], and is stronger than strict balance, see Definition 2.1-(1).

Our approach relies on partition diagrams introduced in Section 4 and later used in Section 5 to arrange the partitions involved in cumulant expressions into a planar representation. This planar representation method has been introduced in [ŁR92] to study the behaviour of variance of subgraph counts in the Erdős-Rényi model, and is extended here to derive cumulant growth rates of all orders in the random-connection model. In this paper, it is used in Section 6 to identify the partition diagrams that play a leading role in cumulant expressions, and yields cumulant growth rates in Proposition 7.1.

We proceed as follows. After recalling necessary preliminaries on the random-connection model and balanced graphs in Section 2, we present our main results in Section 3. Sections 4 and 5 focus the planar diagram representation of cumulants, and Section 6 identifies the leading diagrams appearing in cumulant expressions. Finally, growth rates for cumulants are derived in Section 7. SageMath and R codes used for the computation of convex hulls and for partition counting are listed in Appendices A and B.

2 Preliminaries and notation

Random-connection model

Given $\lambda > 0$ and μ a diffuse sigma-finite measure on ^{d}, $d \geq 1$, we consider a Poisson point process η on $\mathbb{R}^d$ with intensity measure of the form $\lambda \cdot \mu$, which can be almost surely written as

$$\eta = \sum_{i=1}^{\tau} \delta_{X_i}$$

under the probability measure $\mathbb{P}_\lambda$, see [LP18, Corollary 6.5], where τ is a $\mathbb{N} \cup \{\infty\}-$valued random variable, δ_x denotes the Dirac measure at $x \in \mathbb{R}^d$, and $X_1, X_2, \ldots$ are random elements in $\mathbb{R}^d$. For fixed $m \geq 0$, and $y_1, \ldots, y_m \in \mathbb{R}^d$, we consider the point process $\eta \cup \{y_1, \ldots, y_m\}$ on $\mathbb{R}^d$ defined by the union of η and $y_1, \ldots, y_m$.

Given $\varphi : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ a symmetric measurable function and $c_\lambda \in (0,1)$, we also let $\varphi_\lambda := c_\lambda \cdot \varphi$, $\lambda > 0$, denote the connection function of the random-connection model $G_\varphi(\eta \cup \{y_1, \ldots, y_m\})$. The random-connection model is a random graph denoted by $G_\varphi(\eta \cup \{y_1, \ldots, y_m\})$, with vertex set $\eta \cup \{y_1, \ldots, y_m\}$, such that any two distinct vertices $x, y \in \eta \cup \{y_1, \ldots, y_m\}$ are independently connected by an edge with probability $\varphi_\lambda(x, y)$.

Balanced graphs

In what follows, for any two graphs G_1, G_2, we write $G_1 \simeq G_2$ when G_1 is isomorphic to G_2. We also let $v(G) := |V_G| \geq 2$ and $e(G) := |E_G|$ be the number of vertices and the number of edges of any graph G.

Definition 2.1. *[ŁR92], [JŁR00, pages 64-65]*

1) A graph G is balanced if

$$\frac{e(H)}{v(H)} \leq \frac{e(G)}{v(G)}, \qquad H \subset G, \tag{2.1}$$

and strictly balanced if (2.1) *holds as a strict inequality for all $H \subsetneq G$.*

2) A graph G is strongly balanced if

$$\frac{e(H)}{v(H)-1} \leq \frac{e(G)}{v(G)-1}, \qquad H \subset G, \tag{2.2}$$

and strictly strongly balanced if (2.2) *holds as a strict inequality for all $H \subsetneq G$.*

3) A graph G is K_2-balanced if

$$\frac{e(H)-1}{v(H)-2} \leq \frac{e(G)-1}{v(G)-2}, \qquad H \subset G,\ v(H) \geq 3, \tag{2.3}$$

and strictly K_2-balanced if (2.3) *holds as a strict inequality for all $H \subsetneq G$.*

Remark 2.2. From [ŁR92] we have the following statements.

i) Cycles and complete graphs are strictly K_2-balanced.

ii) Trees are K_2-balanced, but not strictly K_2-balanced.

iii) K_2-balanced graph are strongly balanced, except for the unions of disjoint edges, also called matchings.

iv) Strongly balanced graphs are strictly balanced.

Graphs with endpoints

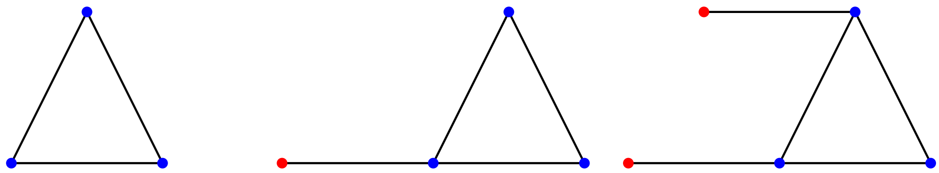
Throughout this paper, we consider a connected graph G satisfying the following conditions.

Assumption 2.3. *Given $r \geq 2$ and $m \geq 0$, we consider a connected graph $G = (V_G, E_G)$ with edge set E_G and vertex set $V_G = \{1, \ldots, r+m\}$, such that*

i) the subgraph induced by G on $\{1, \ldots, r\}$ is connected, and

ii) the endpoint vertices $r+1, \ldots, r+m$ are not adjacent to each other in G,

where Condition (ii) *is void and $V_G = \{1, \ldots, r\}$ in case $m = 0$.*



(a) No endpoint, $m = 0$. (b) One endpoint (rooted graph). (c) Two endpoints, $m = 2$.

Figure 1: Examples of triangles with endpoints, $r = 3$.

In the sequel, we denote $[n] := \{1, \ldots, n\}$ for any $n \geq 1$, and write $V_G = [r + m]$.

Definition 2.4. *We let*

$$a_m(G) := \max_{i \in [r]} |A_i| \tag{2.4}$$

denote the maximum number of endpoint connections to any vertex in $[r]$*, where*

$$A_i := \{j \in \{r+1, \ldots, r+m\} \ : \ \{i, j\} \in E_G\}, \tag{2.5}$$

is the neighborhood of vertex $i \in [r]$ *within the set* $\{r+1, \ldots, r+m\}$ *of endpoints.*

We note that

$$a_m(G) \leq m, \quad a_0(G) = 0, \quad \text{and} \quad a_1(G) = 1.$$

In what follows, our main results will hold under the balance condition

$$\frac{e(H) - a_m(G)}{v(H) - m - 1} \leq \frac{e(G) - a_m(G)}{r - 1}, \qquad H \subset G,\ v(H) \geq m + 2. \tag{2.6}$$

We also note the following points.

Remark 2.5. a) When $r = 2$ and $m \geq 0$, Condition (2.6) is satisfied by all connected graphs.

b) When $r \geq 3$ and $m \geq 0$, Condition (2.6) is satisfied by any tree G, if $m = a_m(G)$. Indeed, when G is a tree and H is a subgraph of G, we have

$$\frac{e(G) - a_m(G)}{r - 1} = \frac{r - 1 + m - a_m(G)}{r - 1} = 1 + \frac{m - a_m(G)}{r - 1}$$

and

$$\frac{e(H) - a_m(G)}{v(H) - m - 1} \leq \frac{v(H) - 1 - a_m(G)}{v(H) - m - 1} = 1 + \frac{m - a_m(G)}{v(H) - m - 1},$$

hence (2.6) is satisfied if $m = a_m(G)$.

c) When $r \geq 2$ and $m = 0$, (2.6) is the strong balance condition (2.2) since $a_0(G) = 0$.

d) When $r \geq 2$ and $m = 1$, (2.6) is the K_2-balance condition (2.3) since $a_1(G) = 1$. Note that cycles, trees and complete graphs are K_2-balanced by Remark 2.2.

Table 1 presents the counts of (isomorphic) trees ($\mathfrak{t}$) *vs.* graphs ($\mathfrak{g}$) satisfying Condition (2.6) within those ($\mathfrak{a}$) satisfying Assumption 2.3 in the format $\mathfrak{t}/\mathfrak{g}/\mathfrak{a}$, using the R code presented in Appendix B for different values of $r \geq 2$ and $m \geq 0$, with $m + r \leq 8$. We check that when $m \geq 1$ the number of trees satisfying Condition (2.6) depends only on $r \geq 3$, and that only trees can satisfy Condition (2.6) as the number m of endpoints becomes large. The first row ($m = 0$) refers to strongly balanced graphs.

Table 1: Counts $\mathfrak{t}/\mathfrak{g}/\mathfrak{a}$ of graphs G satisfying Condition (2.6) *vs.* Assumption 2.3

m \ r	2	3	4	5	6
0	1/1/1	1/2/2	2/5/6	3/14/21	6/53/112
1	1/2/2	2/6/8	4/20/44	9/106/333	20/893/3771
2	2/4/4	2/6/27	4/26/274	9/176/4071	20/2273/94584
3	2/6/6	2/2/73	4/7/1346	9/27/39159	
4	3/9/9	2/2/171	4/4/5620		
5	3/12/12	2/2/359			
6	4/16/16				

3 Main results

Let $y_1, \dots, y_m \in \mathbb{R}^d$ be fixed endpoints, or terminal nodes, where $m \geq 0$ and by convention we set $\{y_1, \dots, y_m\} = \emptyset$ when $m = 0$. In what follows, we consider the count N_G of subgraphs isomorphic to a given connected graph G in the random-connection model $G_\varphi(\eta \cup \{y_1, \dots, y_m\})$ which includes the fixed endpoints $y_1, \dots, y_m$ as vertices.

Definition 3.1. *Let N_G denote the count of (labelled) subgraphs $H \subset G_\varphi(\eta \cup \{y_1, \dots, y_m\})$ such that there exists a bijection $\psi : [r+m] \to V_H$ satisfying*

$$\{i, j\} \in E_G \quad \textit{iff} \quad \{\psi(i), \psi(j)\} \in E_H$$

for $1 \leq i \neq j \leq r+m$, and $\psi(l) = y_l$, $l = 1, \dots, m$.

We note that the subgraph count N_G can be represented as

$$N_G := \sum_{(\alpha_1, \dots, \alpha_r) \in [\tau]^r_{\neq}} \left(\prod_{\substack{\{i,j\} \in E_G \\ i \in [r],\ j \in [r]}} \mathbf{1}_{\{X_{\alpha_i} \leftrightarrow X_{\alpha_j}\}} \right) \left(\prod_{\substack{\{r+j,i\} \in E_G \\ i \in [r],\ j \in [m]}} \mathbf{1}_{\{y_j \leftrightarrow X_{\alpha_i}\}} \right)$$

where $x \leftrightarrow y$ indicates that $x, y \in \eta \cup \{y_1, \dots, y_m\}$ are connected by an edge in the RCM, and

$$[\tau]^r_{\neq} := \{(i_1, \dots, i_r) \in [\tau]^r : i_k \neq i_j \text{ if } k \neq j\}.$$

When $m = 0$, N_G is the count of graphs isomorphic to the graph G in the Poisson random-connection model $G_\varphi(\eta)$.

Assumption 3.2. *i) When $m = 0$, μ is a finite diffuse measure on ^{d}, and $\varphi : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ is a symmetric measurable function.*

ii) When $m \geq 1$, μ is a diffuse sigma-finite measure on $\mathbb{R}^d$, and $\varphi : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ is a symmetric measurable function that satisfies

$$\sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(x, y)\, \mu(\mathrm{d}y) < \infty. \tag{3.1}$$

iii) When $m \geq 1$, in addition to (ii)*, μ is the Lebesgue measure on ^{d} and $\varphi : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ is a symmetric measurable function which is translation invariant, i.e.*

$$\varphi(x, y) = \varphi(0, y - x), \quad x, y \in \mathbb{R}^d.$$

Under Assumption 3.2-(iii), the integrability condition (3.1) reads

$$\int_{\mathbb{R}^d} \varphi(0, x)\, \mathrm{d}x < \infty. \tag{3.2}$$

Normal approximation

Recall that by e.g. Theorem 1 in [Jan88], any sequence $(X_n)_{n\geq 1}$ of real-valued such that

$$\lim_{n\to\infty} \kappa_m(X_n) = 0, \quad \text{for all } m \geq m_0,$$

for some $m_0 \geq 3$ converges in distribution to the Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, provided that the limits

$$\mu := \lim_{n\to\infty} \kappa_1(X_n) \quad \text{and} \quad \sigma^2 := \lim_{n\to\infty} \kappa_2(X_n)$$

exist. Theorem 3.3, which is a consequence of Propositions 6.1 and 7.2, provides sufficient conditions for the asymptotic vanishing of higher order cumulants in (3.3). This will further enable us to apply the method of cumulant for normal approximation in Theorem 3.5.

Theorem 3.3. *Let G be a connected graph with $V_G = [r+m]$ for $r \geq 2$ and $m \geq 0$, and suppose that Assumptions 3.2-*(i)-(iii) *are satisfied and the balance condition* (2.6) *holds. Then, the cumulant $\kappa_n(\overline{N}_G)$ of order $n \geq 1$ of the normalized subgraph count*

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

satisfies the cumulant bound

$$|\kappa_n(\overline{N}_G)| \leq \frac{n!^r}{\Delta_\lambda^{n-2}}, \tag{3.3}$$

where

$$\Delta_\lambda \asymp \begin{cases} \lambda^{r/2} c_\lambda^{e(G)/2} & \text{if } c_\lambda \lesssim \lambda^{-(r-1)/(e(G)-a_m(G))}, \\ \lambda^{1/2} c_\lambda^{a_m(G)/2} & \text{if } 1 \gtrsim c_\lambda \gtrsim \lambda^{-(r-1)/(e(G)-a_m(G))}, \end{cases} \tag{3.4}$$

as λ tends to infinity. In particular, when G has no endpoints $(m=0)$*, we have, as λ tends to infinity,*

$$\Delta_\lambda \asymp \begin{cases} \lambda^{v(G)/2} c_\lambda^{e(G)/2} & \text{if } c_\lambda \lesssim \lambda^{-(v(G)-1)/e(G)}, \\ \lambda^{1/2} & \text{if } 1 \gtrsim c_\lambda \gtrsim \lambda^{-(v(G)-1)/e(G)}, \end{cases}$$

and (2.6) *becomes the strong balance condition.*

Theorem 3.3 extends Corollaries 6.4 and 6.6 of [LP24] without restriction to the dilute and sparse regimes considered therein. When G is a tree with $v(G) = r$ vertices and no endpoints $(m = 0)$, Theorem 3.3 yields

$$\Delta_\lambda \asymp \begin{cases} \lambda^{v(G)/2} c_\lambda^{(v(G)-1)/2} & \text{if } c_\lambda \lesssim \dfrac{1}{\lambda}, \\ \lambda^{1/2} & \text{if } 1 \gtrsim c_\lambda \gtrsim \dfrac{1}{\lambda}, \end{cases}$$

which recovers Corollaries 6.4 and 6.6-1) of [LP24] as particular cases.

Corollary 3.4 (Normal approximation)**.** *Let G be a connected graph with $V_G = [r+m]$ for $r \geq 2$ and $m \geq 0$, suppose that Assumptions 3.2-*(i)-(iii) *are satisfied and the balance condition* (2.6) *holds. In addition, assume that*

$$1 \gtrsim c_\lambda \gg \lambda^{-\min(r/e(G), 1/a_m(G))}. \tag{3.5}$$

Then, the normalized subgraph count $\overline{N}_G$ *converges in distribution to a standard normal random variable as* λ *tends to infinity. In particular, when* $m = 0$ *or* $m = 1$*, Condition* (3.5) *reduces to*

$$1 \gtrsim c_\lambda \gg \lambda^{-r/e(G)},$$

i.e. $1 \gtrsim c_\lambda \gg \lambda^{-v(G)/e(G)}$ *when* G *has no endpoints* $(m = 0)$.

Proof. It suffices to note that under (3.5), in both cases

i) $1 \gtrsim c_\lambda \gg \lambda^{-r/e(G)}$ if $a_m(G)r < e(G)$, and

ii) $1 \gtrsim c_\lambda \gg \lambda^{-1/a_m(G)}$ if $a_m(G)r \geq e(G)$,

we have $\lim_{\lambda\to\infty} \Delta_\lambda = \infty$ in Theorem 3.3, and to apply Theorem 1 in [Jan88]. □

When $a_m(G) = 0$, and in particular if $m = 0$, we have $1/a_m(G) = \infty$, and therefore

$$\min(r/e(G), 1/a_m(G)) = r/e(G)$$

in (3.5). When $c_\lambda \asymp \lambda^{-\alpha}$, the normal approximation result of Corollary 3.4 holds provided that

$$0 < \alpha < \min\left(\frac{r}{e(G)}, \frac{1}{a_m(G)}\right). \tag{3.6}$$

When $m = 0$ or $m = 1$, Condition (3.6) is equivalent to

$$0 < \alpha < \frac{r}{e(G)},$$

which also reads

$$0 < \alpha < \frac{r}{e(G)} = \frac{v(G)}{e(G)}$$

in the absence of endpoints $(m = 0)$.

Theorem 3.5 follows from Theorem 3.3 and the "main lemmas" in Chapter 2 of [SS91] and in [DJS22]. When $m = 0$, the Kolmogorov rate in (3.7) with $c_\lambda \gtrsim \lambda^{-(v(G)-1)/e(G)}$ is consistent with the rate in Corollary 4.6 of [PS20] in the Erdős-Rényi model, up to an additional power $1/(2r-1)$.

Theorem 3.5 (Normal approximation)**.** *Let* G *be a connected graph with* $V_G = [r+m]$ *for* $r \geq 2$ *and* $m \geq 0$*. Suppose that Assumptions 3.2*-(i)-(iii) *are satisfied, that the balance condition* (2.6) *holds, that*

$$1 \gtrsim c_\lambda \gg \lambda^{-\min(r/e(G), 1/a_m(G))},$$

and let Δ_λ *be defined in* (3.4).

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

$$\sup_{x\in\mathbb{R}} |\mathbb{P}_\lambda(\overline{N}_G \leq x) - \Phi(x)| \leq \frac{C}{(\Delta_\lambda)^{1/(2r-1)}}, \tag{3.7}$$

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

ii) (Moderate deviation principle, [DE13, Theorem 1.1] and [DJS22, Theorem 3.1]). Let $(a_\lambda)_{\lambda>0}$ *be a function of* λ *tending to infinity as* λ *tends to infinity, and such that*

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

Then, $(a_\lambda^{-1}\overline{N}_G)_{\lambda>0}$ *satisfies a moderate deviation principle with speed* a_λ^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 λ,

$$\mathbb{P}_\lambda(|\overline{N}_G| \geq x) \leq 2\exp\left(-\frac{1}{4}\min\left(\frac{x^2}{2^r}, (x\Delta_\lambda)^{1/r}\right)\right).$$

iv) (Normal approximation with Cramér corrections, [SS91, Lemma 2.3] and [DJS22, Theorem 2.3]). There exists a constant $c > 0$ such that for all $\lambda \geq 1$ and $x \in (0, c(\Delta_\lambda)^{1/(2r-1)})$ we have

$$\frac{\mathbb{P}_\lambda(\overline{N}_G \geq x)}{1-\Phi(x)} = \left(1 + O\left(\frac{x+1}{(\Delta_\lambda)^{1/(2r-1)}}\right)\right)\exp\left(\widetilde{L}(x)\right)$$

and

$$\frac{\mathbb{P}_\lambda(\overline{N}_G \leq -x)}{\Phi(-x)} = \left(1 + O\left(\frac{x+1}{(\Delta_\lambda)^{1/(2r-1)}}\right)\right)\exp\left(\widetilde{L}(-x)\right),$$

where $\widetilde{L}(x) := (x/c)^3(\Delta_\lambda)^{-3/(2r-1)}\theta$, for some $\theta \in [-1,1]$ depending on $x \in (0, c(\Delta_\lambda)^{1/(2r-1)})$.

Poisson approximation

In what follows, $|\mathrm{Aut}_\bullet(G)|$ stands for the number of automorphisms of the induced subgraph $H \subset G$ with $V_H = [r]$. For example, taking G in Figure 3, the induced graph of G is a path H with vertex set $[4]$ and edge set $E_H = \{\{1,2\},\{1,3\},\{3,4\}\}$, which gives $|\mathrm{Aut}_\bullet(G)| = 2$. The condition $a_m(G)r \leq e(G)$ in the Poisson limit Theorem 3.6 always holds when $m = 0$, and it holds for trees, cycles and complete graphs when $m = 1$.

Theorem 3.6 (Poisson approximation)**.** *Let G be a connected graph with $V_G = [r+m]$ for $r \geq 2$ and $m \geq 0$, suppose that Assumptions 3.2-*(i)-(ii) *are satisfied. If the balance condition* (2.6) *holds together with $a_m(G)r \leq e(G)$ and*

$$\lim_{\lambda\to\infty} \lambda c_\lambda^{e(G)/r} = c > 0,$$

then the subgraph count $\widehat{N}_G := N_G/|\mathrm{Aut}_\bullet(G)|$ converges in distribution to a Poisson random variable with mean

$$\mu_\varphi := \frac{c^r}{|\mathrm{Aut}_\bullet(G)|}\int_{(\mathbb{R}^d)^r}\prod_{\{i,j\}\in E_G}\varphi(x_i,x_j)\,\mu(\mathrm{d}x_1)\cdots\mu(\mathrm{d}x_r),$$

where $x_{r+i} := y_i$ for $i = 1,\ldots,m$.

The proof of Theorem 3.6 is postponed to the end of Section 7.

By the first and second moment methods, see [JŁR00, Page 54], we have

$$\frac{(\mathbb{E}_\lambda[X])^2}{\mathbb{E}_\lambda[X^2]} \leq \mathbb{P}_\lambda(X > 0) \leq \mathbb{E}_\lambda[X] \tag{3.8}$$

for any non-negative integer-valued random variable X, which yields the following threshold result for subgraph containment as a consequence of Corollary 7.2.

Theorem 3.7 (Threshold for subgraph containment)**.** *Let G be a connected graph with $V_G = [r+m]$ for $r \geq 2$ and $m \geq 0$, and suppose that Assumptions 3.2-*(i)-(ii) *are satisfied. If the balance condition* (2.6) *holds together with $a_m(G)r \leq e(G)$, then we have the following threshold results:*

a) $\lim_{\lambda\to\infty}\mathbb{P}_\lambda(N_G = 0) = 1$ if $c_\lambda \ll \lambda^{-r/e(G)}$, *b)* $\lim_{\lambda\to\infty} \mathbb{P}_\lambda(N_G = 0) = e^{-\nu_\varphi}$ *if* $c_\lambda \sim \lambda^{-r/e(G)}$*, with*

$$\nu_\varphi := \frac{1}{|\mathrm{Aut}_\bullet(G)|} \int_{(\mathbb{R}^d)^r} \prod_{\{i,j\}\in E_G} \varphi(x_i, x_j)\, \mu(\mathrm{d}x_1)\cdots\mu(\mathrm{d}x_r),$$

where $x_{r+i} := y_i$ *for* $i = 1, \ldots, m$*,*

c) $\lim_{\lambda\to\infty} \mathbb{P}_\lambda(N_G = 0) = 0$ *if* $1 \gtrsim c_\lambda \gg \lambda^{-r/e(G)}$.

Proof. (a) Since $\mathbb{E}_\lambda[N_G] \asymp \lambda^r c_\lambda^{e(G)}$, if $c_\lambda \ll \lambda^{-r/e(G)}$, we know that $\lim_{\lambda\to\infty} \mathbb{E}_\lambda[N_G] = 0$, and we conclude by the first moment method in (3.8).

(b) Since $c_\lambda \sim \lambda^{-r/e(G)}$, we have $c = \lim_{\lambda\to\infty} \lambda c_\lambda^{e(G)/r} = 1$, and we conclude by Theorem 3.6.

(c) From Corollary 7.2 we know that if $c_\lambda \gg \lambda^{-(r-1)/(e(G)-a_m(G))}$, then

$$\kappa_2(N_G) \asymp \lambda^{2r-1} c_\lambda^{2e(G)-a_m(G)},$$

and if $c_\lambda \ll \lambda^{-(r-1)/(e(G)-a_m(G))}$, then

$$\kappa_2(N_G) \asymp \lambda^r c_\lambda^{e(G)}.$$

If $a_m(G) \le e(G)/r$ and $c_\lambda \gg \lambda^{-r/e(G)}$, then we have

$$\frac{(\mathbb{E}_\lambda[N_G])^2}{\mathbb{E}_\lambda[(N_G)^2]} \asymp \frac{\lambda^{2r} c_\lambda^{2e(G)}}{\lambda^{2r} c_\lambda^{2e(G)} + \kappa_2(N_G)} \asymp \frac{1}{1 + \left(\lambda^r c_\lambda^{e(G)}\right)^{-1} \vee \left(\lambda c_\lambda^{a_m(G)}\right)^{-1}} \to 1, \tag{3.9}$$

and we conclude by the second moment method in (3.8). □

4 Diagram representation of cumulants

This section introduces the diagram framework used for the expansion of cumulants as sums over partitions in Proposition 4.1 below.

We start with basic notation on set partitions, see e.g. [PT11], [LP24]. For any finite set b, we let $\Pi(b)$ denote the collection of set partitions of b. For two set partitions $\rho_1, \rho_2 \in \Pi(b)$, we say ρ_1 is coarser than ρ_2 (i.e. ρ_2 is finer than ρ_1), and we write it as $\rho_2 \preceq \rho_1$, 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 coarsest partition of b, whereas $\widehat{0}$ stands for the partition made of singletons.

Definition 4.1. *Given* $r \ge 2$ *and* $n \ge 1$ *we let* π *denote the partition* $\pi := \{\pi_1, \ldots, \pi_n\} \in \Pi([n]\times[r])$ *of* $[n]\times[r]$ *defined as*

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

a) *A partition* $\rho \in \Pi([n]\times[r])$ *is said to be non-flat if* $\rho \wedge \pi = \widehat{0}$*, and connected if* $\rho \vee \pi = \widehat{1}$.

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

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

denote the set of all connected and non-flat partitions of $[n]\times[r]$*, for* $n, r \ge 1$.

Example 4.2. Figure 2 presents an example of non-flat connected partition in $\mathrm{CNF}(3,4)$.

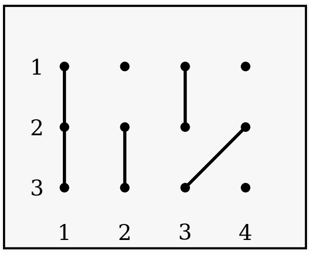


Figure 2: Non-flat connected partition of $[3] \times [4]$.

We also note the following lemma.

Lemma 4.3. *[LP24, 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.*

In [LP24], a graphical diagram language has been designed for the cumulant representation of subgraph counts in the random-connection model, and extended to the case of subgraphs containing fixed endpoints in c.f. [LP25b].

Definition 4.4. *[LP25b] Let G be a connected graph of order $r+m$ satisfying Assumption 2.3, and let $\rho = \{b_1, \ldots, b_{|\rho|}\} \in \Pi([n] \times [r])$, $n \geq 1$, be a partition of $[n] \times [r]$. We let $\bar{\rho}_G$ denote the connected multigraph built on $[m] \cup ([n] \times [r])$, which is constructed as follows.*

1. *For all $i \in [n]$ and $j_1, j_2 \in [r]$, $j_1 \neq j_2$, an edge links (i, j_1) to (i, j_2) iff $(j_1, j_2) \in E_G$;*
2. *For all $k \in [m]$, $i \in [n]$ and $j \in [r]$, an edge links (k) to (i, j) iff $(r+k, j) \in E_G$;*
3. *For all $i \in [|\rho|]$, all elements in the same block b_i are regarded as one vertex.*

In addition, we let ρ_G be the graph constructed from the multigraph $\bar{\rho}_G$ by replacing multiple edges with simple edges in $\bar{\rho}_G$.

In what follows, the blocks of any given partition $\rho = \{b_1, \ldots, b_{|\rho|}\}$ in $\Pi([n] \times [r])$ are ordered along the lexicographic order on $[n] \times [r]$, by ordering the blocks according to their smallest elements. From the above construction, the vertices of ρ_G originate from terminal nodes $[m]$ and blocks $b_1, \ldots, b_{|\rho|}$ of ρ. We can further denote the vertex set of the graph ρ_G as $V(\rho_G) := [|\rho| + m]$ according to the rule that the m terminal nodes follows $b_1, \ldots, b_{|\rho|}$ in order.

See also [Kho08] for a diagram representation used for lines and cycles in the Erdős-Rényi model, and [FGY23] for a graphical representation defined for the U-statistics of determinantal point processes.

Example 4.5. Consider $\rho \in \Pi([3] \times [4])$ as in Figure 2, with

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

and let G be the connected graph with vertex set $V_G = [5]$, represented in Figure 3.

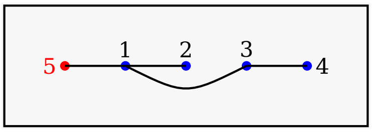
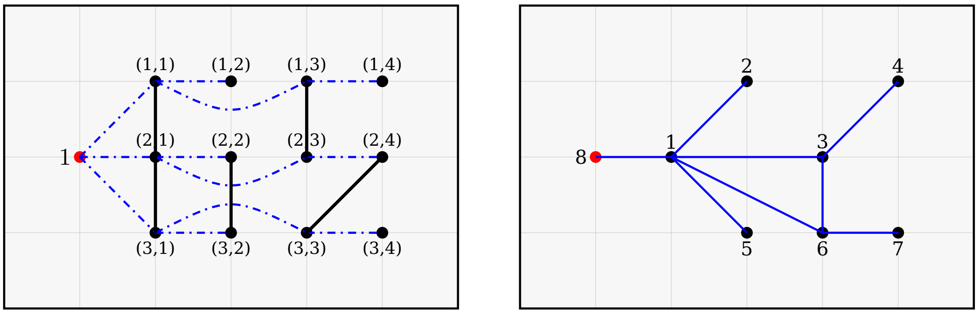


Figure 3: Connected graph G on five vertices including one endpoint, with $r = 4$ and $m = 1$.

Figure 4 presents the multigraph $\bar{\rho}_G$ and corresponding graph ρ_G.



(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 $n = 3$, $r = 4$, and $m = 1$.

The above framework allows us to state moment and cumulant formulas for the subgraph counts N_G.

Definition 4.6. *For $\rho = \{b_1, \dots, b_{|\rho|}\} \in \Pi([n] \times [r])$ and $j \in [m]$, we denote by*

$$\mathcal{A}_j^\rho := \{k \in [|\rho|] : \ \exists (s,i) \in b_k \text{ s.t. } \{i, r+j\} \in E_G\} \tag{4.1}$$

the neighborhood of the vertex $(|\rho| + j)$ in the graph ρ_G.

From [LP25b], we have the following moment and cumulant representation for N_G.

Proposition 4.1. Let $n \geq 1$. Then, the n-th moments and n-th cumulants of N_G admit the expressions

$$\mathbb{E}_\lambda[(N_G)^n] = \sum_{\substack{\rho \in \Pi([n]\times[r]) \\ \rho \wedge \pi = \widehat{0} \\ (\text{non-flat})}} F_\lambda(\rho) \quad \text{and} \quad \kappa_n(N_G) = \sum_{\substack{\rho \in \Pi_{\widehat{1}}([n]\times[r]) \\ \rho \wedge \pi = \widehat{0} \\ (\text{non-flat connected})}} F_\lambda(\rho), \tag{4.2}$$

where $F_\lambda(\rho)$, $\rho \in \Pi([n] \times [r])$, is defined as

$$F_\lambda(\rho) := \lambda^{|\rho|} \int_{(\mathbb{R}^d)^{|\rho|}} \prod_{\substack{1 \leq j \leq m \\ i \in \mathcal{A}_j^\rho}} \varphi_\lambda(x_i, y_j) \prod_{\substack{1 \leq k < l \leq |\rho| \\ \{k,l\} \in E(\rho_G)}} \varphi_\lambda(x_k, x_l) \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_{|\rho|}). \tag{4.3}$$

5 Planar representation of partition diagrams

In this section, we introduce a planar representation that will allow us to determine the leading partition diagrams in the moment and cumulant expressions of Proposition 4.1.

Definition 5.1. *Let G be a connected graph with $V_G = [r+m]$, for some $r \geq 2$ and $m \geq 0$. For $n \geq 2$, we let*

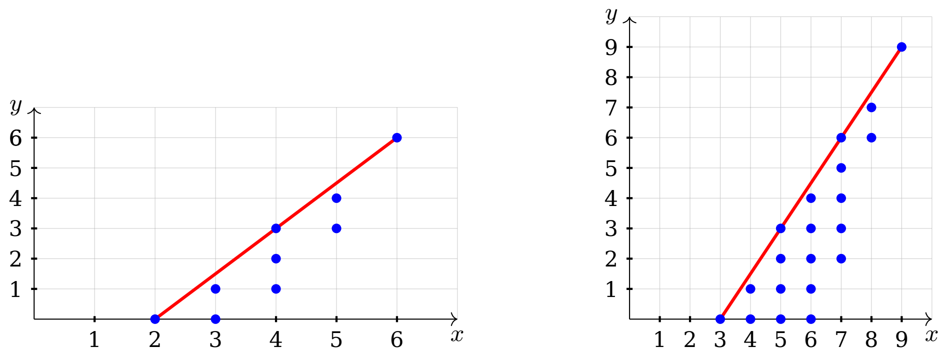
$$\Sigma_n(G, m) := \big\{(x(\rho_G), y(\rho_G)) := (nr + m - v(\rho_G), ne(G) - e(\rho_G)) : \ \rho \in \mathrm{CNF}(n, r)\big\},$$

where, for every partition $\rho \in \Pi([n] \times [r])$, ρ_G is the graph associated to ρ by Definition 4.4.

Example 5.2. Let $G = C_3$ be a triangle with no endpoint, i.e. $r = 3$ and $m = 0$. We have

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

see Figure 5.



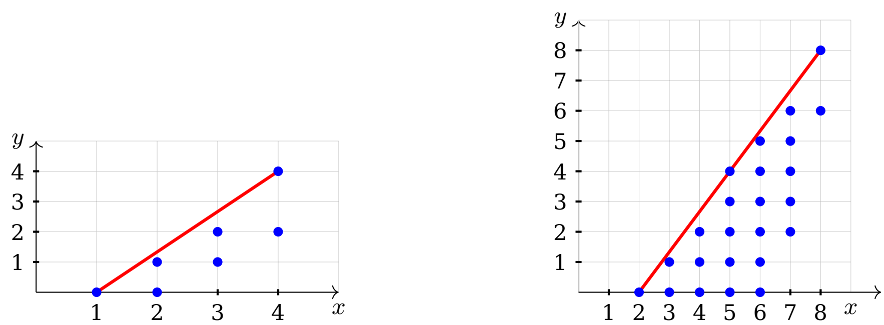
(a) Subgraph plot of $\Sigma_3(C_3, 0)$. (b) Subgraph plot of $\Sigma_4(C_3, 0)$.

Figure 5: Set $\Sigma_n(C_3, 0)$ and upper boundary of its convex hull (in red) for $n = 3, 4$.

Figure 5 and the following ones can be plotted after loading the SageMath code presented in the appendix and running the following commands.

```
G = [[1,2],[2,3],[3,1]]; EP = []; SG3=convexhull(3,G,EP); SG4=convexhull(4,G,EP)
Polyhedron(SG3).plot(color = "pink")+point(SG3,color = "blue",size=20)
Polyhedron(SG4).plot(color = "pink")+point(SG4,color = "blue",size=20)
```

Example 5.3. Let $G = C_4$ be a 4-cycle with no endpoint, i.e. $r = 4$ and $m = 0$. Here, $\Sigma_2(C_4, 0)$ and $\Sigma_3(C_4, 0)$ are plotted in Figure 6.



(a) Subgraph plot of $\Sigma_2(C_4, 0)$. (b) Subgraph plot of $\Sigma_3(C_4, 0)$.

Figure 6: Set $\Sigma_n(C_4, 0)$ and upper boundary of its convex hull (in red) for $n = 2, 3$.

Figure 6 can be plotted via the following commands, after loading the SageMath code listed in the appendix.

```
G = [[1,2],[2,3],[3,4],[4,1]]; EP = []; SG2=convexhull(2,G,EP); SG3=convexhull(3,G,EP)
Polyhedron(SG2).plot(color = "pink")+point(SG2,color = "blue",size=20)
Polyhedron(SG3).plot(color = "pink")+point(SG3,color = "blue",size=20)
```

Example 5.4. Let $G = C_4$ be a rooted 4-cycle with one endpoint, i.e. $r = 3$ and $m = 1$, see Figure 7.

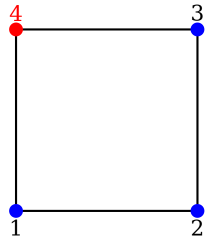
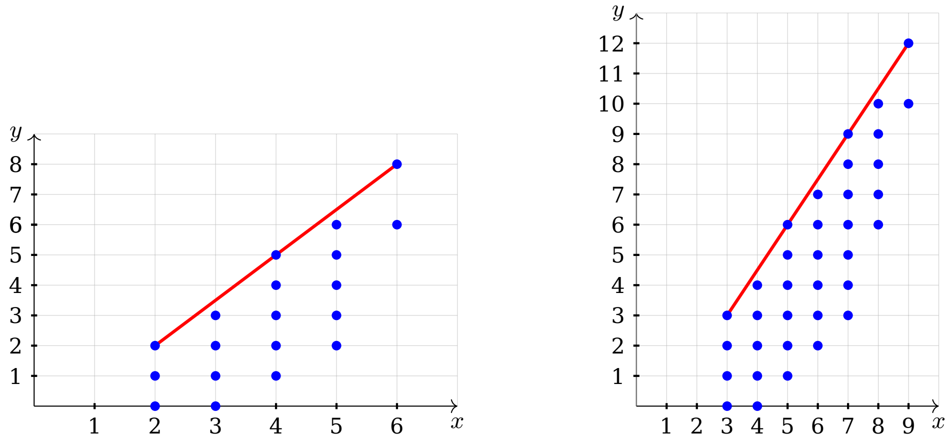


Figure 7: Connected graph $G = C_4$ with $r = 3$ and $m = 1$.

The sets $\Sigma_3(C_4, 1)$ and $\Sigma_4(C_4, 1)$ are plotted in Figure 8.



(a) Subgraph plot of $\Sigma_3(C_4, 1)$.

(b) Subgraph plot of $\Sigma_4(C_4, 1)$.

Figure 8: Set $\Sigma_n(C_4, 1)$ and upper boundary of its convex hull (in red) for $n = 3, 4$.

Figure 8 can be plotted via the following commands.

```
G = [[1,2],[2,3]]; EP = [[1,3]]; SG3=convexhull(3,G,EP); SG4=convexhull(4,G,EP)
Polyhedron(SG3).plot(color = "pink")+point(SG3,color = "blue",size=20)
Polyhedron(SG4).plot(color = "pink")+point(SG4,color = "blue",size=20)
```

Figures 5-8 also show an upper boundary plotted in red, which is characterized in the next definition.

Definition 5.5. *We let* $\widehat{\Sigma}_n(G, m)$ *denote the upper boundary of the convex hull of* $\Sigma_n(G, m)$*, with*

$$\widehat{\Sigma}_n(G, m) \cap \Sigma_n(G, m) = \{(x_i, z_i) : i = 0, 1, \ldots, l\}, \tag{5.1}$$

for some $l \geq 1$*, where*

$$n - 1 = x_0 < x_1 < \cdots < x_l = (n-1)r < x_{l+1} := +\infty.$$

In Definition 5.1, $x(\rho_G) = nr + m - v(\rho_G) = nr - |\rho|$ stands for the number of vertices being removed in the process of graph contraction. Later on, in Definition 5.5 x_0 and x_l will be used to denote the minimum and the maximum number of vertices being removed. Because for all $\rho \in \mathrm{CNF}(n, r)$,

$$r \leq |\rho| \leq n(r-1) + 1,$$

we have $x_0 = n-1$ and $x_l = (n-1)r$. We note that for any point (x_i, z_i) in $\widehat{\Sigma}_n(G,m) \cap \Sigma_n(G,m)$, $i \in \{0, 1, \ldots, l\}$, there exists a connected non-flat partition $\rho_i \in \mathrm{CNF}(n,r)$ such that the associated graph $\rho_{i,G}$ satisfies

$$v(\rho_{i,G}) = nr + m - x_i \quad \text{and} \quad e(\rho_{i,G}) = ne(G) - z_i. \tag{5.2}$$

We note that the upper boundary $\widehat{\Sigma}_n(G,m)$ starts at $(x_0, y_0) := (n-1, (n-1)a_m(G))$ and ends at $((n-1)r, (n-1)e(G))$, where $a_m(G)$ is defined in (2.4).

We also recall the following lemma from [LP24, Lemma 2.8], in which the maximality of connected non-flat partitions refers to maximizing the number of blocks, see also Proposition 6.1 in [ST24].

Lemma 5.6. *$a)$ The cardinality of the set $\mathrm{CNF}(n,r)$ of connected non-flat partitions of $[n] \times [r]$ satisfies*

$$|\mathrm{CNF}(n,r)| \le n!^r r!^{n-1}, \qquad n, r \ge 1. \tag{5.3}$$

$b)$ Let $\mathcal{M}(n,r)$ denote the set of maximal connected non-flat partitions of $[n] \times [r]$. Then, each element of $\mathcal{M}(n,r)$ has precisely $(n-1)r+1$ blocks, and we have

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

with the bounds

$$((r-1)r)^{n-1}(n-1)! \le |\mathcal{M}(n,r)| \le ((r-1)r)^{n-1} n!, \quad n \ge 1, \ r \ge 2. \tag{5.4}$$

6 Leading diagrams

Based on the convex hull of $\Sigma_n(G,m)$ given in Definition 5.1, in this section we identify the dominant asymptotic order and the leading contribution appearing in the expression (4.2) of $\kappa_n(N_G)$, $n \ge 1$, which is key to the derivation of normal approximation results via the cumulant method.

Definition 6.1. *Given G a connected graph with $V_G = [r+m]$, a diagram $\rho \in \mathrm{CNF}(n,r)$, $n \ge 1$, is said to be a leading diagram for a given $(c_\lambda)_{\lambda > 0}$ if, for every $\sigma \in \mathrm{CNF}(n,r)$ satisfies that*

$$\lambda^{v(\sigma_G)} c_\lambda^{e(\sigma_G)} = O\big(\lambda^{v(\rho_G)} c_\lambda^{e(\rho_G)}\big), \qquad \text{as } \lambda \to \infty.$$

The characterization of leading diagrams will use the following definition.

Definition 6.2. *Let $\rho \in \mathrm{CNF}(n,r)$ and $i_\rho \in \{0, \ldots, l\}$ be such that*

$$x_{i_\rho} \le x(\rho_G) < x_{i_\rho+1},$$

where $x(\rho_G)$ is given in Definition 5.1. Using the notation (5.1), we define

$$\theta_-(\rho_G) := \begin{cases} +\infty, & i_\rho = 0, \\ \dfrac{z_{i_\rho} - z_{i_\rho-1}}{x_{i_\rho} - x_{i_\rho-1}}, & 1 \le i_\rho \le l, \end{cases} \quad \text{and} \quad \theta_+(\rho_G) := \begin{cases} \dfrac{z_{i_\rho+1} - z_{i_\rho}}{x_{i_\rho+1} - x_{i_\rho}}, & 0 \le i_\rho < l, \\ 0, & i_\rho = l. \end{cases}$$

We note the inequality

$$\theta_-(\rho_G) \ge \theta_+(\rho_G),$$

which holds because $\widehat{\Sigma}_n(G,m)$ is the upper boundary of the convex hull of $\Sigma_n(G,m)$. We also say that a diagram $\rho \in \mathrm{CNF}(n,r)$ lies on the boundary $\widehat{\Sigma}_n(G,m)$ if the point $(x(\rho_G), y(\rho_G))$ does. Lemma 6.3 states that leading diagrams can only lie on the upper boundary $\widehat{\Sigma}_n(G,m)$.

Lemma 6.3. *Let G be a connected graph with $V_G = [r+m]$, $r \geq 2$, $m \geq 0$. Let $n \geq 2$ and assume that $\lim_{\lambda\to\infty} c_\lambda = 0$.*

1) Every leading diagram $\rho \in \mathrm{CNF}(n,r)$ lies on the upper boundary $\widehat{\Sigma}_n(G,m)$, i.e.

$$(nr + m - v(\rho_G), ne(G) - e(\rho_G)) \in \widehat{\Sigma}_n(G,m).$$

2) If a diagram $\rho \in \mathrm{CNF}(n,r)$ lies on the upper boundary $\widehat{\Sigma}_n(G,m)$ and

$$\lambda c_\lambda^{\theta_-(\rho_G)} = O(1) \text{ and } \lambda c_\lambda^{\theta_+(\rho_G)} = \Omega(1),$$

then ρ is a leading diagram.

Proof. 1) Suppose that $\rho \in \mathrm{CNF}(n,r)$ does not lie on the boundary $\widehat{\Sigma}_n(G,m)$, i.e.

$$(x(\rho_G), y(\rho_G)) := (nr + m - v(\rho_G), ne(G) - e(\rho_G)) \in \Sigma_n(G,m) \backslash \widehat{\Sigma}_n(G,m).$$

Using (5.1), if $x(\rho_G) = x_l$, we know that $y(\rho_G) < z_l$, as $(x(\rho_G), y(\rho_G))$ is not on the upper boundary $\widehat{\Sigma}_n(G,m)$. Therefore, we have

$$\lambda^{v(\rho_G) - v(\rho_{l,G})} c_\lambda^{e(\rho_G) - e(\rho_{l,G})} = c_\lambda^{z_l - y(\rho_G)} \ll 1,$$

and ρ cannot be a leading diagram.

If $x(\rho_G) < x_l$, we choose $i_\rho \in \{0, \ldots, l-1\}$ such that $x_{i_\rho} \leq x(\rho_G) < x_{i_\rho+1}$. Since $\widehat{\Sigma}_n(G,m)$ is the upper boundary of a convex hull, by (5.2) if $i_\rho < l$ we have

$$\nu := \frac{z_{i_\rho+1} - y(\rho_G)}{x_{i_\rho+1} - x(\rho_G)} > \theta_+(\rho_G) = \frac{z_{i_\rho+1} - z_{i_\rho}}{x_{i_\rho+1} - x_{i_\rho}} = \frac{e(\rho_{i_\rho,G}) - e(\rho_{i_\rho+1,G})}{v(\rho_{i_\rho,G}) - v(\rho_{i_\rho+1,G})},$$

i.e. $\lambda c_\lambda^\nu \ll \lambda c_\lambda^{\theta_+(\rho_G)}$ because $\lim_{\lambda\to\infty} c_\lambda = 0$, and we consider three cases.

i) If $1 \ll \lambda c_\lambda^\nu$, we have

$$\begin{aligned}
\lambda^{v(\rho_G) - v(\rho_{i_\rho+1,G})} c_\lambda^{e(\rho_G) - e(\rho_{i_\rho+1,G})} &= (\lambda c_\lambda^\nu)^{v(\rho_G) - v(\rho_{i_\rho+1,G})} \\
&\ll \left(\lambda c_\lambda^{\theta_+(\rho_G)}\right)^{v(\rho_G) - v(\rho_{i_\rho+1,G})} \\
&\leq \left(\lambda c_\lambda^{\theta_+(\rho_G)}\right)^{v(\rho_{i_\rho,G}) - v(\rho_{i_\rho+1,G})} \\
&= \lambda^{v(\rho_{i_\rho,G}) - v(\rho_{i_\rho+1,G})} c_\lambda^{e(\rho_{i_\rho,G}) - e(\rho_{i_\rho+1,G})},
\end{aligned}$$

which implies $\lambda^{v(\rho_G)} c_\lambda^{e(\rho_G)} \ll \lambda^{v(\rho_{i_\rho,G})} c_\lambda^{e(\rho_{i_\rho,G})}$ as $\lim_{\lambda\to\infty} c_\lambda = 0$, hence the diagram ρ is not leading.

ii) If $\lim_{\lambda\to\infty} \lambda c_\lambda^\nu = c > 0$, we have

$$\begin{aligned}
\lambda^{v(\rho_G) - v(\rho_{i_\rho+1,G})} c_\lambda^{e(\rho_G) - e(\rho_{i_\rho+1,G})} &= (\lambda c_\lambda^\nu)^{v(\rho_G) - v(\rho_{i_\rho+1,G})} \\
&\ll \left(\lambda c_\lambda^{\theta_+(\rho_G)}\right)^{v(\rho_G) - v(\rho_{i_\rho+1,G})} \\
&\asymp \left(\lambda c_\lambda^{\theta_+(\rho_G)}\right)^{v(\rho_{i_\rho,G}) - v(\rho_{i_\rho+1,G})} \\
&= \lambda^{v(\rho_{i_\rho,G}) - v(\rho_{i_\rho+1,G})} c_\lambda^{e(\rho_{i_\rho,G}) - e(\rho_{i_\rho+1,G})},
\end{aligned}$$

and we conclude as above.

iii) If $\lambda c_\lambda^\nu \ll 1$, we have

$$\lambda^{v(\rho_G)-v(\rho_{i_\rho+1,G})} c_\lambda^{e(\rho_G)-e(\rho_{i_\rho+1,G})} \ll 1,$$

hence

$$\lambda^{v(\rho_G)} c_\lambda^{e(\rho_G)} \ll \lambda^{v(\rho_{i_\rho+1,G})} c_\lambda^{e(\rho_{i_\rho+1,G})}.$$

As a consequence of the above, we find

$$\lambda^{v(\rho_G)} c_\lambda^{e(\rho_G)} \ll \lambda^{v(\rho_{i_\rho,G})} c_\lambda^{e(\rho_{i_\rho,G})} \text{ or } \lambda^{v(\rho_G)} c_\lambda^{e(\rho_G)} \ll \lambda^{v(\rho_{i_\rho+1,G})} c_\lambda^{e(\rho_{i_\rho+1,G})},$$

hence ρ is not a leading diagram.

2) Suppose that ρ does lie on the boundary $\widehat{\Sigma}_n(G,m)$. Then, there exists $i_\rho \in \{0,\dots,l\}$ such that $(x(\rho_G), y(\rho_G)) = (x_{i_\rho}, z_{i_\rho})$, and it holds that

$$\lambda c_\lambda^{\theta_-(\rho_G)} = O(1) \text{ and } \lambda c_\lambda^{\theta_+(\rho_G)} = \Omega(1).$$

i) If $j < i_\rho$, then $x_j - x_{i_\rho} < 0$ and

$$\frac{z_{i_\rho} - z_j}{x_{i_\rho} - x_j} \geq \frac{z_{i_\rho} - z_{i_\rho-1}}{x_{i_\rho} - x_{i_\rho-1}} = \theta_-(\rho_G).$$

Hence,

$$\lambda c_\lambda^{(z_{i_\rho}-z_j)/(x_{i_\rho}-x_j)} = O\big(\lambda c_\lambda^{\theta_-(\rho_G)}\big) = O(1). \tag{6.1}$$

Now, since

$$\frac{\lambda^{nr+m-x_{i_\rho}} c_\lambda^{ne(G)-y_{i_\rho}}}{\lambda^{nr+m-x_j} c_\lambda^{ne(G)-z_j}} = \lambda^{x_j-x_{i_\rho}} c_\lambda^{z_j-z_{i_\rho}} = \big(\lambda c_\lambda^{(z_j-z_{i_\rho})/(x_j-x_{i_\rho})}\big)^{x_j-x_{i_\rho}}, \tag{6.2}$$

we find

$$\frac{\lambda^{v(\rho_G)} c_\lambda^{e(\rho_G)}}{\lambda^{v(\rho_{j,G})} c_\lambda^{e(\rho_{j,G})}} = \Omega(1).$$

ii) If $j > i_\rho$, then $x_j - x_{i_\rho} > 0$ and

$$\frac{z_j - z_{i_\rho}}{x_j - x_{i_\rho}} \leq \theta_+(\rho_G),$$

therefore

$$\lambda c_\lambda^{(z_{i_\rho}-z_j)/(x_{i_\rho}-x_j)} = \Omega\big(\lambda c_\lambda^{\theta_+(\rho_G)}\big) = \Omega(1), \tag{6.3}$$

and (6.2) shows that

$$\frac{\lambda^{v(\rho_G)} c_\lambda^{e(\rho_G)}}{\lambda^{v(\rho_{j,G})} c_\lambda^{e(\rho_{j,G})}} = \Omega(1).$$

This ensures that ρ is a leading diagram. □

Proposition 6.1 provides a sufficient condition ensuring that the upper boundary $\widehat{\Sigma}_n(G,m)$ is a line segment, a property used in the proof of Corollary 7.2.

Proposition 6.1. Let G be a connected graph with $V_G = [r+m]$, $r \geq 2$, $m \geq 0$, such that the balance condition (2.6) holds. Then, the upper boundary $\widehat{\Sigma}_n(G,m)$ is a line segment for all $n \geq 1$.

Proof. To present the result in a more compact form, we will first show that the requirement that the upper boundary $\widehat{\Sigma}_n(G,m)$ is a line segment is equivalent to

$$\frac{e(G)-a_m(G)}{r-1} \le \frac{e(\rho_G)-a_m(G)}{v(\rho_G)-m-1}, \qquad \rho \in \mathrm{CNF}(n,r). \tag{6.4}$$

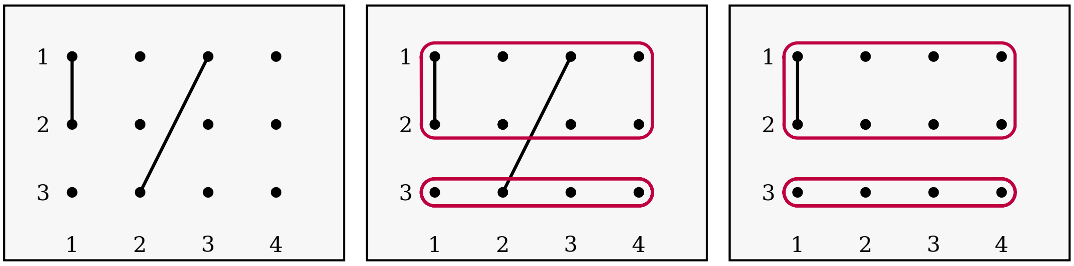
Because the upper boundary $\widehat{\Sigma}_n(G,m)$ starts at $(x_0,z_0) := (n-1,(n-1)a_m(G))$ and ends at $(x_l,z_l) := ((n-1)r,(n-1)e(G))$, the requirement that the upper boundary $\widehat{\Sigma}_n(G,m)$ is a line segment is equivalent to that for any $(x,z) \in \Sigma_n(G,m)$,

$$\frac{z_l - z}{x_l - x} \ge \frac{z_l - z_0}{x_l - x_0} = \frac{e(G)-a_m(G)}{r-1}. \tag{6.5}$$

Considering the Definition 5.1, we obtain that the requirement itself is further equivalent to for any diagram $\rho \in \mathrm{CNF}(n,r)$,

$$\frac{e(\rho_G)-e(G)}{v(\rho_G)-m-r} \ge \frac{e(G)-a_m(G)}{r-1}, \tag{6.6}$$

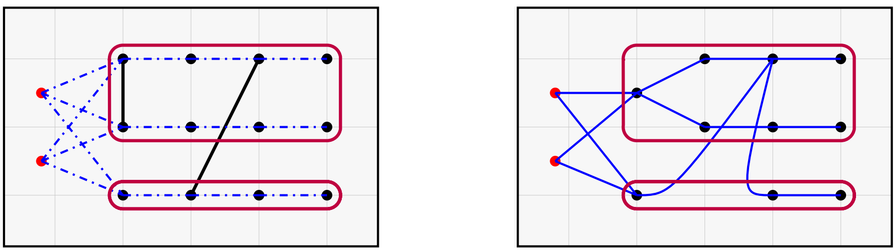
and, after reorganizing, the above inequality becomes equivalent to (6.4). It remains to show that the balance condition (2.6) ensures that (6.4) is satisfied. Here, we apply an induction argument to see this. When $n=1$, the claim is trivial as the only element in $\mathrm{CNF}(1,r)$ is isomorphic to G. Suppose now that (6.4) holds up to the rank $n \ge 1$. Let $\rho \in \mathrm{CNF}(n+1,r)$ be a non-flat connected partition of $[n+1]\times[r]$ with associated graph ρ_G. By Lemma 4.3, up to reordering of $\{1,\dots,n+1\}$ there exists a partition $\sigma \in \mathrm{CNF}(n,r)$ obtained by restriction of ρ to $[n]\times[r]$. Let σ' denote the partition obtained by restriction of ρ to $\{n+1\}\times[r]$, see Figure 9.



(a) Partition ρ of $[n+1]\times[r]$. (b) Splitting of $[n+1]\times[r]$. (c) Partitions σ and σ'.

Figure 9: Splitting of a partition ρ into σ and σ' with $n=3$ and $r=4$.

Given a graph G with r vertices, let ρ_G denote the graph with vertex set $V(\rho_G)$ built on ρ as in Definition 4.4, see Figure 10 for an example with G a graph on $r+m=6$ vertices including two endpoints.



(a) Vertex set $V(\rho_G)$. (b) Graph ρ_G.

Figure 10: Splitting of ρ_G into σ_G and σ'_G with $n=3$, $r=4$ and two endpoints $m=2$.

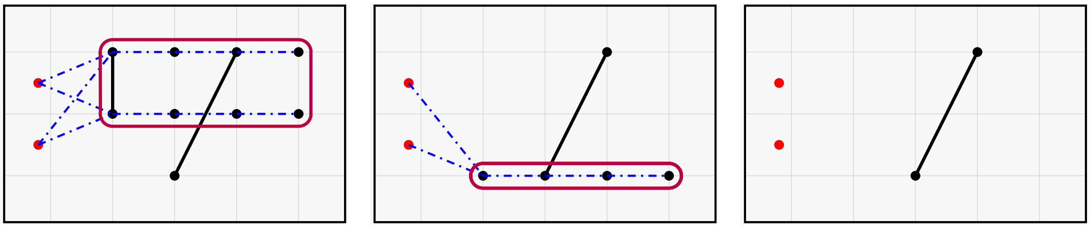
Next, we consider the graphs σ_G and σ'_G obtained from Definition 4.4 on the vertex sets

$$V(\sigma_G) := \{b \in \rho \ : \ b \cap (\pi_1 \cup \cdots \cup \pi_n) \neq \emptyset\} \cup [m]$$

and

$$V(\sigma'_G) := \{b \in \rho \ : \ b \cap \pi_{n+1} \neq \emptyset\} \cup [m],$$

with $\sigma'_G \simeq G$ because ρ is non-flat, see Figure 11.



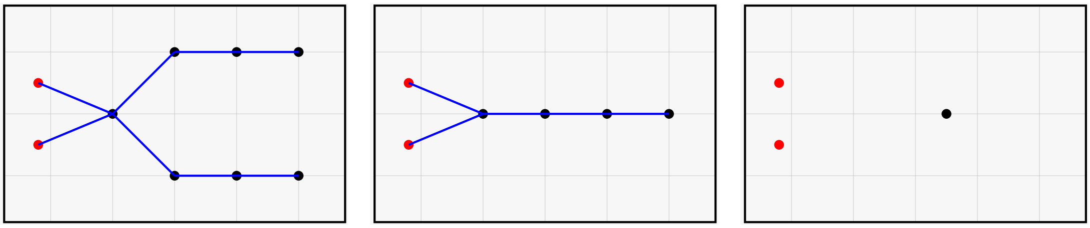
(a) Vertex set $V(\sigma_G)$.

(b) Vertex set $V(\sigma'_G)$.

(c) Vertex set $V(\sigma''_G)$.

Figure 11: Splitting of $V(\rho_G)$ into $V(\sigma_G), V(\sigma'_G)$ with $n = 3$, $r = 4$ and two endpoints $m = 2$.

Let now σ''_G denote the graph induced by ρ_G on $V(\sigma''_G) := V(\sigma_G) \cap V(\sigma'_G)$, see Figure 12.



(a) Graph σ_G.

(b) Graph σ'_G.

(c) Graph σ''_G.

Figure 12: Splitting of ρ_G into σ_G and σ'_G with $n = 3$, $r = 4$ and two endpoints $m = 2$.

Then, σ''_G contains m endpoints in addition to at least one non-endpoint vertex due to the connectedness of ρ, hence we have $v(\sigma''_G) \geq m + 1$. Since $\sigma''_G \subset \sigma'_G$ and $v(\sigma''_G) \geq m + 1$, by the balance condition (2.6) we have

$$\frac{e(\sigma''_G) - a_m(G)}{v(\sigma''_G) - m - 1} \leq \frac{e(G) - a_m(G)}{r - 1},$$

with the convention $0/0 = 0$. Hence, by the induction hypothesis (6.4) applied at the rank $n \geq 1$ to σ_G, we have

$$\begin{aligned}
\frac{e(\rho_G) - a_m(G)}{v(\rho_G) - m - 1} &= \frac{(e(\sigma_G) - a_m(G)) + (e(\sigma'_G) - a_m(G)) - (e(\sigma''_G) - a_m(G))}{v(\sigma_G) + v(\sigma'_G) - v(\sigma''_G) - m - 1} \\
&= \frac{(e(\sigma_G) - a_m(G)) + (e(G) - a_m(G)) - (e(\sigma''_G) - a_m(G))}{v(\sigma_G) + v(G) - v(\sigma''_G) - m - 1} \\
&\geq \frac{(v(\sigma_G) - m - 1)\frac{e(G)-a_m(G)}{r-1} + (e(G) - a_m(G)) - (v(\sigma''_G) - m - 1)\frac{e(G)-a_m(G)}{r-1}}{v(\sigma_G) + v(G) - v(\sigma''_G) - m - 1} \\
&= \frac{e(G) - a_m(G)}{r - 1}.
\end{aligned}$$

□

The balance condition (2.6) turns out to be necessary in order to ensure the upper boundary $\widehat{\Sigma}_n(G,m)$ to be a line segment, as shown in the following counterexample.

Counterexample 6.2. Consider the graph G of Figure 13, which is not strongly balanced, with $r = 4$ and $m = 0$.

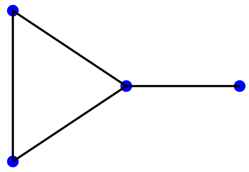
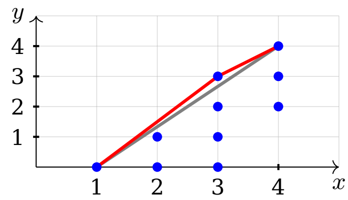
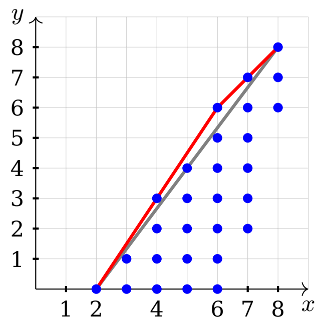


Figure 13: Not strongly balanced graph G.

Figure 14 shows that the upper boundary $\widehat{\Sigma}_n(G,m)$ is not a line segment for $n = 2$ and $n = 3$. $\widehat{\Sigma}_2(G,0)$ and $\widehat{\Sigma}_3(G,0)$.



(a) Subgraph plot of $\Sigma_2(G,0)$ with $\widehat{\Sigma}_2(G,0)$.



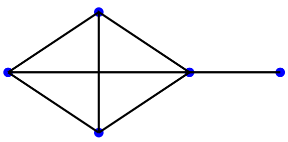
(b) Subgraph plot of $\Sigma_3(G,0)$ and $\widehat{\Sigma}_3(G,0)$.

Figure 14: Set $\Sigma_n(G,0)$ and upper boundary of its convex hull (in red) for $n = 2,3$.

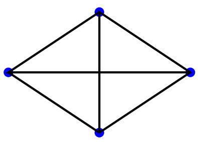
Figure 14 can be plotted via the following commands.

```
G = [[1,2],[2,3],[3,4],[1,3]]; EP = []; SG2=convexhull(2,G,EP); SG3=convexhull(3,G,EP)
Polyhedron(SG2).plot(color = "pink")+point(SG2,color = "blue",size=20)
Polyhedron(SG3).plot(color = "pink")+point(SG3,color = "blue",size=20)
```

Remark 6.4. We note that the balance condition (2.6) is not necessary for asymptotic normality of normalized subgraph counts. Consider the graph G in Figure 15 for example, where $r = 5, e(G) = 7$, $m = a_m(G) = 0$, which is not strongly balanced, and not even balanced.



(a) graph G.



(b) subgraph $H \subset G$.

Figure 15: A (not balanced) graph G and subgraph H.

Since (2.6) is not satisfied, the upper boundary $\widehat{\Sigma}_n(G,m)$ is not a line segment, which leads to more potential candidates for leading diagram, beyond $\lambda^{1+(r-1)n}c_\lambda^{ne(G)-(n-1)a_m(G)}$

and $\lambda^r c_\lambda^{e(G)}$. Precisely, we have

$$\begin{aligned}\kappa_n(N_G) &\asymp \max\left\{\lambda^{nr-(n-1)}c_\lambda^{ne(G)}, \lambda^{nr-(n-1)v(H)}c_\lambda^{ne(G)-(n-1)e(H)}, \lambda^r c_\lambda^{e(G)}\right\} \\ &= \max\left\{\lambda^{4n+1}c_\lambda^{7n}, \lambda^{n+4}c_\lambda^{n+6}, \lambda^5 c_\lambda^7\right\} \\ &= \begin{cases} \lambda^{4n+1}c_\lambda^{7n} & \text{if } c_\lambda \gg \lambda^{-1/2}, \\ \lambda^{n+4}c_\lambda^{n+6} & \text{if } \lambda^{-1} \ll c_\lambda \lesssim \lambda^{-1/2}, \\ \lambda^5 c_\lambda^7 & \text{if } c_\lambda \lesssim \lambda^{-1}. \end{cases}\end{aligned}$$

Therefore, when $c_\lambda \gg \lambda^{-2/3}$ we have $\kappa_n(\overline{N}_G) \to 0$, $n \geq 3$, as λ tends to infinity, which implies asymptotic normality of $\overline{N}_G$ by Theorem 1 in [Jan88]. On the other hand, when $c_\lambda \lesssim \lambda^{-1}$ we have $\kappa_n(N_G) \to 0$, $n \geq 1$, therefore N_G does not have a Poisson limit.

7 Cumulant growth rates for subgraph counts

Under Assumptions 3.2-(i)-(ii), $F_\lambda(\rho)$ defined in (4.3) satisfies

$$F_\lambda(\rho) \asymp \lambda^{|\rho|}c_\lambda^{e(\rho_G)}. \tag{7.1}$$

In this section, we investigate the asymptotic behaviour of the cumulants $\kappa_n(N_G)$ in (4.2) as $c_\lambda \to 0$ and the intensity λ tends to infinity, by identifying the leading diagrams $\rho \in \mathrm{CNF}(n,r)$ which, from (4.2) and Definition 6.1, satisfy

$$\kappa_n(N_G) \asymp \lambda^{v(\rho_G)-m}c_\lambda^{e(\rho_G)}. \tag{7.2}$$

Proposition 7.1. Let G be a connected graph with $V_G = [r+m]$ for $r \geq 2$ and m endpoints, $m \geq 0$. Suppose that Assumptions 3.2-(i)-(ii) are satisfied and that the upper boundary $\widehat{\Sigma}_n(G,m)$ is a line segment linking $(n-1,(n-1)a_m(G))$ to $((n-1)r,(n-1)e(G))$.

a) If $1 \gtrsim c_\lambda \gtrsim \lambda^{-(r-1)/(e(G)-a_m(G))}$, we have

$$\kappa_n(N_G) \asymp \lambda^{1+(r-1)n}c_\lambda^{ne(G)-(n-1)a_m(G)}, \quad n \geq 2. \tag{7.3}$$

b) If $c_\lambda \asymp \lambda^{-(r-1)/(e(G)-a_m(G))}$, we have

$$\kappa_n(N_G) \asymp \lambda c_\lambda^{a_m(G)}, \quad n \geq 2. \tag{7.4}$$

c) If $c_\lambda \lesssim \lambda^{-(r-1)/(e(G)-a_m(G))}$, we find

$$\kappa_n(N_G) \asymp \lambda^r c_\lambda^{e(G)}, \quad n \geq 2. \tag{7.5}$$

Proof. As in (5.1), we write

$$\widehat{\Sigma}_n(G,m) \cap \Sigma_n(G,m) = \{(x_0,z_0),(x_1,z_1),\dots,(x_l,z_l)\},$$

with $x_0 := n-1 < x_1 < \cdots < x_l := (n-1)r$. According to Definition 5.1, we can find a corresponding partition $\rho_{i,G} \in \mathrm{CNF}(n,r)$ such that

$$v(\rho_{i,G}) = nr + m - x_i, \quad e(\rho_{i,G}) = ne(G) - z_i.$$

Also, we write the (connected non-flat) set partition associated with $\rho_{i,G}$ as ρ_i, and from (7.1) we obtain that each ρ_i contributes

$$F_\lambda(\rho_i) \asymp \lambda^{v(\rho_{i,G})-m}c_\lambda^{e(\rho_{i,G})} \asymp \lambda^{nr-x_i}c_\lambda^{ne(G)-z_i}. \tag{7.6}$$

Because the upper boundary is a line segment with endpoints $(n-1,(n-1)a_m(G))$ and $((n-1)r,(n-1)e(G))$, the slope of this line segment is $\theta := (e(G)-a_m(G))/(r-1)$. By Lemma 6.3-(1) the leading diagram ρ must lie on $\widehat{\Sigma}_n(G,m)$.

a) For any $j = 1, \dots, l$, by (5.2) we have

$$\begin{aligned}\frac{\lambda^{v(\rho_{0,G})-m} c_\lambda^{e(\rho_{0,G})}}{\lambda^{v(\rho_{j,G})-m} c_\lambda^{e(\rho_{j,G})}} &= \frac{\lambda^{1+(r-1)n} c_\lambda^{ne(G)-(n-1)a_m(G)}}{\lambda^{nr-x_j} c_\lambda^{ne(G)-z_j}} \\ &= \frac{\lambda^{nr-x_0} c_\lambda^{ne(G)-z_0}}{\lambda^{nr-x_j} c_\lambda^{ne(G)-z_j}} \\ &= \lambda^{x_j-x_0} c_\lambda^{z_j-z_0} \\ &= \lambda^{x_j-x_0} c_\lambda^{\theta(x_j-x_0)} \\ &= \left(\lambda c_\lambda^{(e(G)-a_m(G))/(r-1)}\right)^{x_j-x_0},\end{aligned}$$

hence if $\lambda c_\lambda^{(e(G)-a_m(G))/(r-1)} \gtrsim 1$, we find

$$\frac{\lambda^{v(\rho_{0,G})-m} c_\lambda^{e(\rho_{0,G})}}{\lambda^{v(\rho_{j,G})-m} c_\lambda^{e(\rho_{j,G})}} \gtrsim 1,$$

therefore any ρ_G such that

$$(v(\rho_G)-m, e(\rho_G)) = (v(\rho_{0,G})-m, e(\rho_{0,G})) = (1+(r-1)n, ne(G)-(n-1)a_m(G))$$

is a leading diagram, and this yields (7.3) by (7.2).

b) If $\lambda c_\lambda^{(e(G)-a_m(G))/(r-1)} \asymp 1$ then any diagram ρ_i, $i = 0, 1, \dots, l$ on the segment $\widehat{\Sigma}_n(G, m)$ is a leading diagram by Lemma 6.3-(2). Furthermore, by choosing $j = l$ with

$$(v(\rho_G)-m, e(\rho_G)) = (v(\rho_{l,G})-m, e(\rho_{l,G})) = (r, e(G)),$$

we find that (7.2) yields (7.4), i.e.

$$\kappa_n(N_G) \asymp \lambda^r c_\lambda^{e(G)} \asymp \lambda c_\lambda^{a_m(G)}.$$

c) For any $j = 0, \dots, l-1$, by (5.2) we have

$$\begin{aligned}\frac{\lambda^{v(\rho_{l,G})-m} c_\lambda^{e(\rho_{l,G})}}{\lambda^{v(\rho_{j,G})-m} c_\lambda^{e(\rho_{j,G})}} &= \frac{\lambda^{r} c_\lambda^{e(G)}}{\lambda^{nr-x_j} c_\lambda^{ne(G)-z_j}} \\ &= \frac{\lambda^{nr-x_l} c_\lambda^{ne(G)-z_l}}{\lambda^{nr-x_j} c_\lambda^{ne(G)-z_j}} \\ &= \lambda^{x_j-x_l} c_\lambda^{z_j-z_l} \\ &= \lambda^{x_j-x_l} c_\lambda^{\theta(x_j-x_l)} \\ &= \left(\lambda c_\lambda^{(e(G)-a_m(G))/(r-1)}\right)^{x_j-x_l},\end{aligned}$$

hence if $\lambda c_\lambda^{(e(G)-a_m(G))/(r-1)} \lesssim 1$, we find

$$\frac{\lambda^{v(\rho_{l,G})-m} c_\lambda^{e(\rho_{l,G})}}{\lambda^{v(\rho_{j,G})-m} c_\lambda^{e(\rho_{j,G})}} \gtrsim 1.$$

Therefore, any ρ_G such that

$$(v(\rho_G)-m, e(\rho_G)) = (v(\rho_{l,G})-m, e(\rho_{l,G})) = (r, e(G))$$

is a leading diagram, and this yields (7.5) by (7.2). $\square$

We note from Lemma 6.3 and Proposition 6.1 that as long as a connected graph G satisfies the balance condition (2.6), the leading asymptotic order in the expression (4.2) of $\kappa_n(N_G)$ is fully determined by either the maximal or the minimal connected non-flat partition. Here, maximality, resp. minimality, of partitions refers to the maximality, resp. minimality, of their block counts. As a consequence of Remark 2.5, we have the following.

Remark 7.1. When $m = 1$ with $a_1(G) = 1$, Proposition 7.1 holds for trees, cycles and complete graphs as they are all K_2-balanced and the balance condition (2.6) coincides with the K_2-balance condition (2.3).

By Propositions 6.1 and 7.1, we have the following result.

Corollary 7.2. *Let G be a connected graph with $V_G = [r + m]$, $r \geq 2$, and m endpoints, $m \geq 0$. Suppose that Assumptions 3.2-*(i)-(ii) *and the balance condition* (2.6) *are satisfied. Then, the cumulant $\kappa_n(N_G)$ of order $n \geq 1$ of the subgraph count N_G satisfies the following.*

a) If $1 \gtrsim c_\lambda \gtrsim \lambda^{-(r-1)/(e(G)-a_m(G))}$, we have

$$\kappa_n(N_G) \asymp \lambda^{1+(r-1)n} c_\lambda^{ne(G)-(n-1)a_m(G)}.$$

b) If $c_\lambda \asymp \lambda^{-(r-1)/(e(G)-a_m(G))}$, we have

$$\kappa_n(N_G) \asymp \lambda c_\lambda^{a_m(G)}.$$

c) If $c_\lambda \lesssim \lambda^{-(r-1)/(e(G)-a_m(G))}$, we find

$$\kappa_n(N_G) \asymp \lambda^r c_\lambda^{e(G)}.$$

In addition, by Remark 2.5-a) we have the following consequence of Corollary 7.2.

Corollary 7.3. *Let G be a strongly balanced connected graph with $v(G) = r$ vertices, $r \geq 2$, and no endpoints, i.e. $m = 0$, and suppose that Assumptions 3.2-*(i)-(ii) *are satisfied.*

a) If $1 \gtrsim c_\lambda \gtrsim \lambda^{-(v(G)-1)/e(G)}$, we have

$$\kappa_n(N_G) \asymp \lambda^{1+(v(G)-1)n} c_\lambda^{ne(G)}.$$

b) If $c_\lambda \asymp \lambda^{-(v(G)-1)/e(G)}$, we have

$$\kappa_n(N_G) \asymp \lambda.$$

c) If $c_\lambda \lesssim \lambda^{-(v(G)-1)/e(G)}$, we find

$$\kappa_n(N_G) \asymp \lambda^{v(G)} c_\lambda^{e(G)}.$$

Proposition 7.2 deals with the cumulant growth of normalized subgraph counts, for use in normal approximation. In the particular case of $(r+1)$-hop counting with $c_\lambda = 1$ and $m = 2$, (7.7) is consistent with the normalized cumulant bound (8.2) in [Pri24].

Proposition 7.2. Let G be a connected graph with $V_G = [r + m]$ for $r \geq 2$ and $m \geq 0$. Suppose that Assumptions 3.2-(i)-(iii) are satisfied and that the upper boundary $\widehat{\Sigma}_n(G, m)$ is a line segment linking $(n-1, (n-1)a_m(G))$ to $((n-1)r, (n-1)e(G))$. Denoting by

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

the normalized subgraph count, we have

$$|\kappa_n(\overline{N}_G)| \le \frac{n!^r}{\Delta_\lambda^{n-2}}, \quad n \ge 3, \tag{7.7}$$

where

$$\Delta_\lambda \asymp \begin{cases} \lambda^{1/2} c_\lambda^{a_m(G)/2} & \text{if } 1 \gtrsim c_\lambda \gtrsim \lambda^{-(r-1)/(e(G)-a_m(G))}, \\ \lambda^{(e(G)-ra_m(G))/(2(e(G)-a_m(G)))} & \text{if } c_\lambda \asymp \lambda^{-(r-1)/(e(G)-a_m(G))}, \\ \lambda^{r/2} c_\lambda^{e(G)/2} & \text{if } \lambda^{-r/e(G)} \ll c_\lambda \lesssim \lambda^{-(r-1)/(e(G)-a_m(G))}. \end{cases} \tag{7.8}$$

Proof. We start by assuming that $m \ge 1$. We only focus on the case when $n \ge 3$, as cases $n = 1, 2$ are trivial.

a) When $1 \gtrsim c_\lambda \gtrsim \lambda^{-(r-1)/(e(G)-a_m(G))}$, since the balance condition (2.6) holds, from Proposition 7.1-a), we know that the leading diagrams belong to the set $\mathcal{M}(n,r)$ of maximal connected non-flat partitions of $[n] \times [r]$, see Lemma 5.6. Therefore, given (4.2) and (5.3), we can bound $\kappa_n(N_G)$ from above

$$\begin{aligned} \kappa_n(N_G) &\le |\mathrm{CNF}(n,r)| \lambda^{1+(r-1)n} c_\lambda^{ne(G)-(n-1)a_m(G)} C_{1,n} \\ &\le n!^r r!^{n-1} \lambda^{1+(r-1)n} c_\lambda^{ne(G)-(n-1)a_m(G)} C_{1,n} \end{aligned} \tag{7.9}$$

where

$$C_{1,n} := \max_{\rho \in \mathcal{M}(n,r)} \int_{(\mathbb{R}^d)^{|\rho|}} \prod_{\substack{1 \le j \le m \\ i \in \mathcal{A}_j^\rho}} \varphi(x_i, y_j) \prod_{\substack{1 \le k < l \le |\rho| \\ \{k,l\} \in E(\rho_G)}} \varphi(x_k, x_l) \, \mathrm{d}x_1 \cdots \mathrm{d}x_{|\rho|}. \tag{7.10}$$

Because the function $\varphi : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ is symmetric and translation invariant, we can further bound $C_{1,n}$ as follows. From Definition 4.4, we know that for any $\rho \in \mathcal{M}(n,r)$, ρ_G is a connected graph with $V(\rho_G) = [m+1+n(r-1)]$, as $|\rho| = 1+n(r-1)$. Let $\tilde{\rho}_G$ be the subgraph of ρ_G induced by $V(\tilde{\rho}_G) = [n(r-1)+2]$. And we also denote $\bar{\rho}_G$ a spanning tree of $\tilde{\rho}_G$. Therefore,

$$\begin{aligned} &\int_{(\mathbb{R}^d)^{|\rho|}} \prod_{\substack{1 \le j \le m \\ i \in \mathcal{A}_j^\rho}} \varphi(x_i, y_j) \prod_{\substack{1 \le k < l \le |\rho| \\ \{k,l\} \in E(\rho_G)}} \varphi(x_k, x_l) \, \mathrm{d}x_1 \cdots \mathrm{d}x_{|\rho|} \\ &\le \int_{(\mathbb{R}^d)^{|\rho|}} \prod_{i \in \mathcal{A}_1^\rho} \varphi(x_i, y_1) \prod_{\substack{1 \le k < l \le |\rho| \\ \{k,l\} \in E(\tilde{\rho}_G)}} \varphi(x_k, x_l) \, \mathrm{d}x_1 \cdots \mathrm{d}x_{|\rho|} \\ &\le \int_{(\mathbb{R}^d)^{|\rho|}} \prod_{\substack{1 \le k < l \le |\rho|+1 \\ \{k,l\} \in E(\bar{\rho}_G)}} \varphi(x_k, x_l) \, \mathrm{d}x_1 \cdots \mathrm{d}x_{|\rho|} \\ &= \left(\int_{\mathbb{R}^d} \varphi(0,x) \, \mathrm{d}x \right)^{1+n(r-1)}, \end{aligned} \tag{7.11}$$

where $x_{|\rho|+1} := y_1$, and the last equality is obtained by integrating successively on the variables which correspond to leaves of $\bar{\rho}_G$ as in the proofs of e.g. Theorem 7.1 of [LNS21] or Lemma 3.1 of [CT22] since φ is translation invariant by Assumption 3.2-(iii). Hence, $C_{1,n}$ is bounded by $\zeta_\varphi^{1+n(r-1)}$, where

$$\zeta_\varphi := \max\left(1, \int_{\mathbb{R}^d} \varphi(0,x) \, \mathrm{d}x\right).$$

In the other direction, as in (4.2), the cumulants of N_G are written as a summation of some non-negative terms. Therefore, we can bound $\kappa_n(N_G)$ from below, as follows:

$$\begin{aligned}\kappa_n(N_G) &\geq |\mathcal{M}(n,r)|\lambda^{1+(r-1)n}c_\lambda^{ne(G)-(n-1)a_m(G)}C_{2,n}\\ &\geq ((r-1)r)^{n-1}(n-1)!\lambda^{1+(r-1)n}c_\lambda^{ne(G)-(n-1)a_m(G)}C_{2,n},\end{aligned} \tag{7.12}$$

where the last inequality comes from (5.4), and

$$C_{2,n} := \min_{\rho\in\mathcal{M}(n,r)} \int_{(\mathbb{R}^d)^{|\rho|}} \prod_{\substack{1\leq j\leq m\\ i\in\mathcal{A}_j^\rho}} \varphi(x_i,y_j) \prod_{\substack{1\leq k<l\leq|\rho|\\ \{k,l\}\in E(\rho_G)}} \varphi(x_l,x_k)\,\mathrm{d}x_1\cdots\mathrm{d}x_{|\rho|}. \tag{7.13}$$

Combining (7.9) and (7.12), we have, for $n\geq 3$

$$\begin{aligned}\kappa_n(\overline{N}_G) &= \frac{\kappa_n(N_G)}{\kappa_2(N_G)^{n/2}}\\ &\leq \frac{n!^r r!^{n-1}\lambda^{1+(r-1)n}c_\lambda^{ne(G)-(n-1)a_m(G)}C_{1,n}}{\big((r-1)r\lambda^{1+2(r-1)}c_\lambda^{2e(G)-a_m(G)}C_{2,2}\big)^{n/2}}\\ &= n!^r\frac{r!^{n-1}}{((r-1)r)^{n/2}}\big(\lambda c_\lambda^{a_m(G)}\big)^{-(n-2)/2}\frac{C_{1,n}}{C_{2,2}^{n/2}}\\ &\leq n!^r((r-2)!)^{n-1}((r-1)r)^{n/2-1}\big(\lambda c_\lambda^{a_m(G)}\big)^{-(n-2)/2}\frac{\zeta_\varphi^{1+n(r-1)}}{C_{2,2}^{n/2}}\\ &\leq \frac{n!^r}{\Big(C_3\sqrt{\lambda c_\lambda^{a_m(G)}}\Big)^{n-2}},\end{aligned} \tag{7.14}$$

where C_3 is a constant depending on r and φ.

b) When $c_\lambda \lesssim \lambda^{-(r-1)/(e(G)-a_m(G))}$, from Proposition 7.1-$b$), the leading diagrams are $\rho\in\mathrm{CNF}(n,r)$ such that $\rho_G\simeq G$, which allows us to bound $\kappa_n(N_G)$ as follows:

$$\begin{aligned}\kappa_n(N_G) &\leq |\mathrm{CNF}(n,r)|\lambda^r c_\lambda^{e(G)}\int_{(\mathbb{R}^d)^r}\prod_{\substack{1\leq k<l\leq r+m\\ \{k,l\}\in E_G}}\varphi(x_k,x_l)\,\mathrm{d}x_1\cdots\mathrm{d}x_r\\ &\leq n!^r r!^{n-1}\lambda^r c_\lambda^{e(G)}\zeta_\varphi^r,\end{aligned} \tag{7.15}$$

where $x_{r+i} := y_i$ for $i=1,\dots,m$. What's more, we can also bound $\kappa_2(N_G)$ from below

$$\begin{aligned}\kappa_2(N_G) &\geq \lambda^r c_\lambda^{e(G)}\int_{(\mathbb{R}^d)^r}\prod_{\substack{1\leq k<l\leq r+m\\ \{k,l\}\in E_G}}\varphi(x_k,x_l)\,\mathrm{d}x_1\cdots\mathrm{d}x_r\\ &=: \lambda^r c_\lambda^{e(G)}C_4.\end{aligned} \tag{7.16}$$

Combining (7.15) and (7.16), we get

$$\begin{aligned}\kappa_n(\overline{N}_G) &\leq \frac{n!^r r!^{n-1}\lambda^r c_\lambda^{e(G)}\zeta_\varphi^r}{\big(\lambda^r c_\lambda^{e(G)}C_4\big)^{n/2}}\\ &\leq n!^r\big(\lambda^r c_\lambda^{e(G)}\big)^{-(n-2)/2}\frac{r!^{n-1}\zeta_\varphi^{rn}}{C_4^{n/2}}\\ &\leq \frac{n!^r}{\Big(C_5\sqrt{\lambda^r c_\lambda^{e(G)}}\Big)^{n-2}},\end{aligned} \tag{7.17}$$

where C_5 is a constant depending only on r and φ.

When $m = 0$ the above arguments apply by replacing the upper bound (7.11) with $\mu(^d)^{1+(r-1)n}$. □

We are now ready to prove Theorem 3.6.

Proof of Theorem 3.6. Since $\lim_{\lambda\to\infty} \lambda c_\lambda^{e(G)/r} = c > 0$ and $a_m(G)r \leq e(G)$, we have $c_\lambda \lesssim \lambda^{-(r-1)/(e(G)-a_m(G))}$. Hence, as in the proof of Corollary 7.2-(c), by (4.2) we obtain

$$\begin{aligned}\lim_{\lambda\to\infty} \kappa_n(N_G) &= |\mathrm{Aut}_\bullet(G)|^{n-1} \lim_{\lambda\to\infty} \lambda^r \int_{(\mathbb{R}^d)^r} \left(\prod_{\{i,j\}\in E_G} c_\lambda \varphi(x_i, x_j) \right) \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_r) \\ &= |\mathrm{Aut}_\bullet(G)|^{n-1} c^r \int_{(\mathbb{R}^d)^r} \left(\prod_{\{i,j\}\in E_G} \varphi(x_i, x_j) \right) \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_r), \qquad n \geq 1,\end{aligned}$$

since the count of connected non-flat set partitions $\rho \in \mathrm{CNF}([n] \times [r])$ such that $|\rho| = r$ and $e(\rho_G) = e(G)$ is $|\mathrm{Aut}_\bullet(G)|^{n-1}$. We conclude from Theorem 6.14 in [JŁR00]. □

A Convex hull code

The following SageMath code determines the convex hull of $\Sigma_n(G, m)$ and its upper boundary $\widehat{\Sigma}_n(G, m)$, see Figures 5-6 and 14. This code and the following one are available for download at https://github.com/nprivaul/convex-hull.

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

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

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

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

def graphs(G,EP,setpartition,n):
```

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

def convexhull(n,G,EP):
 r=len(set(flatten(G)));m=len(EP)
 cnfp=connectednonflat(n,r)
 L=[]
 le=sum(len(EP[j]) for j in range(len(EP)))
 for setpartition in cnfp:
    rhoG=graphs(G,EP,setpartition,n)
    edgesrhoG = [i for n, i in enumerate(rhoG) if i not in rhoG[:n]]
    vertrhoG = set(flatten(edgesrhoG));
    L.append((n*r-(len(vertrhoG)-m),n*(len(G)+le)-len(edgesrhoG)))
 return sorted(set(L))
```

B Graph counting code

The following R code uses the graph6 and sparse6 formats for undirected graphs and the data files available at https://users.cecs.anu.edu.au/~bdm/data/graphs.html.

```
library(rgraph6); library(igraph); library(matrixStats)
r=3; m=2; graphs=read_file6("graph5c.g6") # m+r=5
# r=5; m=3; graphs=read_file6("graph8c.g6") # m+r=8
count=0; total=0; trees=0; treesam=0; nontreesr=0;
for (mat in graphs) {g=as.undirected(graph_from_adjacency_matrix(mat))
endpoints=combn(1:(m+r),m); lst = c();
for (k in 1:choose(m+r,m)) {
V(g)$color <- c(7, 2)[1 + V(g) %in% endpoints[,k]]
complement=setdiff(c(1:(m+r)),endpoints[,k])
if(sum(degree(subgraph(g,endpoints[,k])))==0 && is.connected(subgraph(g, complement)))
{if (all(sapply(lst, function(gg) !graph.isomorphic.vf2(g,gg)$iso))) {lst=c(lst,list(g));
count=count+1; a=0; exit=0;
for (ii in complement) {a=max(a,sum(g[ii,endpoints[,k]]))}
for (i in (m+2):(m+r)){ for (j in 1:choose(m+r,i)){
h=subgraph(g, combn(1:(m+r),i)[,j])
if (((ecount(g)-a)/(vcount(g)-m-1))<((ecount(h)-a)/(vcount(h)-m-1))) {
exit=1;break;}}
if (exit==1) {break;}}
if (exit==0) {print(g); cat("a =",a,"\n"); total=total+1
trees=trees+is_tree(g)*(a==m); treesam=treesam+is_tree(g)
nontreesr=nontreesr+!is_tree(subgraph(g, complement))
plot(g); print("Working ...");}}}}}
cat("Tree with a<>m count = ",treesam, ";Tree count = ",trees,"out of total =",total,"out of",
     count, "\n");
```

References

[Bar82] A.D. Barbour. Poisson convergence and random graphs. *Math. Proc. Cambridge Philos. Soc.*, 92(2):349–359, 1982. MR0671189

[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. MR4574856

[BHJ92] A.D. Barbour, L. Holst, and S. Janson. *Poisson approximation*, volume 2 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1992. MR1163825

[BJR07] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures Algorithms*, 31(1):3–122, 2007. MR2337396

[BKR89] A.D. Barbour, M. Karoński, and A. Ruciński. A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B*, 47(2):125–145, 1989. MR1047781

[Bol81a] B. Bollobás. Random graphs. In *Combinatorics, Proceedings, Swansea 1981*, volume 52 of *London Mathematical Society Lecture Notes Series*, pages 80–102. Cambridge University Press, 1981. MR0633650

[Bol81b] B. Bollobás. Threshold functions for small subgraphs. *Math. Proc. Cambridge Philos. Soc.*, 90(2):197–206, 1981. MR0620729

[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. MR3528620

[CGR18] M. Coulson, R. E. Gaunt, and G. Reinert. Compound Poisson approximation of subgraph counts in stochastic block models with multiple edges. *Adv. in Appl. Probab.*, 50(3):759–782, 2018. MR3877253

[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. MR4388562

[DE13] H. Döring and P. Eichelsbacher. Moderate deviations via cumulants. *J. Theoret. Probab.*, 26:360–385, 2013. MR3055808

[DF14] L. Devroye and N. Fraiman. Connectivity of inhomogeneous random graphs. *Random Structures Algorithms*, 45(3):408–420, 2014. MR3252923

[DJS22] H. Döring, S. Jansen, and K. Schubert. The method of cumulants for the normal approximation. *Probab. Surv.*, 19:185–270, 2022. MR4408127

[ER59] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959. MR0120167

[ER61] P. Erdős and A. Rényi. On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, 38:343–347, 1961. MR0148055

[FGY23] R. Feng, F. Götze, and D. Yao. Determinantal point processes on spheres: multivariate linear statistics. *arXiv preprint arXiv:2301.09216*, 2023.

[Gil59] E.N. Gilbert. Random graphs. *Ann. Math. Statist*, 30(4):1141–1144, 1959. MR0108839

[GT18] J. Grote and C. Thäle. Gaussian polytopes: a cumulant-based approach. *J. Complexity*, 47:1–41, 2018. MR3804598

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

[Jan88] S. Janson. Normal convergence by higher semiinvariants with applications to sums of dependent random variables and random graphs. *Ann. Probab.*, 16(1):305–312, 1988. MR0920273

[Jan19] S. Jansen. Cluster expansions for Gibbs point processes. *Adv. in Appl. Probab.*, 51(4):1129–1178, 2019. MR4032174

[JŁR00] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000. MR1782847

[KGKLN23] A.P. Kartun-Giles, K. Koufos, X. Lu, and D. Niyato. Two-hop connectivity to the roadside in a VANET under the random connection model. *IEEE Trans. Veh. Technol.*, 72(4):5508–5512, 2023.

[Kho08] O. Khorunzhiy. On connected diagrams and cumulants of Erdős-Rényi matrix models. *Comm. Math. Phys.*, 282:209–238, 2008. MR2415478

[KR83] M. Karoński and A. Ruciński. On the number of strictly balanced subgraphs of a random graph. In *Graph Theory, Proceedings, Lagów, 1981*, volume 1018 of *Lecture Notes in Mathematics*, pages 79–83, Berlin, 1983. Springer. MR0730636

[KR21] G. Kaur and A. Röllin. Higher-order fluctuations in dense random graph models. *Electron. J. Probab.*, 25:1–36, 2021. MR4346670

[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. MR4254476

[Lov12] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012. MR3012035

[LP18] G. Last and M. Penrose. *Lectures on the Poisson Process*, volume 7 of *Institute of Mathematical Statistics Textbooks*. Cambridge University Press, Cambridge, 2018. MR3791470

[LP24] Q. Liu and N. Privault. Normal approximation of subgraph counts in the random-connection model. *Bernoulli*, 30:3224–3250, 2024. MR4779863

[LP25a] Q. Liu and N. Privault. Gaussian fluctuations of generalized U-statistics and subgraph counting in the binomial random-connection model. Preprint arXiv:2505.12338, 33 pages, 2025.

[LP25b] Q. Liu and N. Privault. Graph connectivity with fixed endpoints in the random-connection model. *Probab. Engrg. Inform. Sci.*, 39(3):370–396, 2025.

[ŁR92] T. Łuczak and A. Ruciński. Convex hulls of dense balanced graphs. *J. Comput. Appl. Math.*, 41:205–213, 1992. MR1181721

[LS06] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006. MR2274085

[Mau24] P.A. Maugis. Central limit theorems for local network statistics. *Biometrika*, 111:743–754, 2024. MR4794147

[MR96] R. Meester and R. Roy. *Continuum percolation*, volume 119 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996. MR1409145

[NZZ$^+$11] S. C. Ng, W. Zhang, Y. Zhang, Y. Yang, and G. Mao. Analysis of access and connectivity probabilities in vehicular relay networks. *IEEE J. Sel. Areas Commun.*, 29(1):140–150, 2011.

[Pen91] M. Penrose. On a continuum percolation model. *Adv. in Appl. Probab.*, 23(3):536–556, 1991. MR1122874

[Pen03] M. Penrose. *Random Geometric Graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, 2003. MR1986198

[Pen16] M. Penrose. Connectivity of soft random geometric graphs. *Ann. Appl. Probab.*, 26(2):986–1028, 2016. MR3476631

[Pen18] M. Penrose. Inhomogeneous random graphs, isolated vertices, and Poisson approximation. *J. Appl. Probab.*, 55(1):112–136, 2018. MR3780386

[Pri24] N. Privault. Asymptotic analysis of k-hop connectivity in the 1D unit disk random graph model. *Methodol. Comput. Appl. Probab.*, 26(47), 2024. MR4812491

[PS20] N. Privault and G. Serafin. Normal approximation for sums of discrete U-statistics - application to Kolmogorov bounds in random subgraph counting. *Bernoulli*, 26(1):587–615, 2020. MR4036045

[PT11] G. Peccati and M. Taqqu. *Wiener Chaos: Moments, Cumulants and Diagrams: A survey with Computer Implementation*. Bocconi & Springer Series. Springer, 2011. MR2791919

[Ruc88] A. Ruciński. When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields*, 78:1–10, 1988. MR0940863

[RV86] A. Ruciński and A. Vince. Strongly balanced graphs and random graphs. *J. Graph Theory*, 10(2):251–264, 1986. MR0890231

[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. MR1171883

[ST24] M. Schulte and C. Thäle. Moderate deviations on Poisson chaos. *Electron. J. Probab.*, 29:1–27, 2024. MR4814710

[ZCY+12] W. Zhang, Y. Chen, Y. Yang, X. Wang, Y. Zhang, X. Hong, and G. Mao. Multi-hop connectivity probability in infrastructure-based vehicular networks. *IEEE J. Sel. Areas Commun.*, 30(4):740–747, 2012.

[Zha22] Z.S. Zhang. Berry-Esseen bounds for generalized U-statistics. *Electron. J. Probab.*, 27:1–36, 2022. MR4492985

Acknowledgments. We thank the referees for useful suggestions.