Check for updates

J. Appl. Probab. 1–16 (2026)
*doi:*10.1017/jpr.2026.10090

ASYMPTOTIC NORMALITY FOR TRIANGLE COUNTING IN THE SPARSE β-MODEL

SIANG ZHANG,*
QUNQIANG FENG ,* ** AND
ZHISHUI HU ,* *University of Science and Technology of China*

Abstract

We study the number of triangles T_n in the sparse β-model on n vertices, a random graph model that captures degree heterogeneity in real-world networks. Using the norms of the heterogeneity parameter vector, we first determine the asymptotic mean and variance of T_n. Next, by applying the Malliavin–Stein method, we derive a non-asymptotic upper bound on the Kolmogorov distance between the normalized T_n and the standard normal distribution. Under an additional assumption on degree heterogeneity, we further prove the asymptotic normality for T_n as $n \to \infty$.

Keywords: Random networks; sparsity; degree heterogeneity; subgraph counts; Malliavin–Stein method

2020 Mathematics Subject Classification: Primary 05C80; 90B15
Secondary 60F05

1. Introduction

Subgraph counting is a fundamental problem in the study of random graphs and has been extensively investigated over several decades. A significant body of research has focused on this topic within the framework of classical Erdös–Rényi (ER) random graphs (see, e.g., [6, 38] and references therein). For instance, as the graph size (i.e. the number of vertices) tends to infinity, the necessary and sufficient conditions for the asymptotic normality of the count of any fixed subgraph in the ER random graph model were first established in the seminal work of Ruciński [37]. More recently, this result has been complemented with explicit bounds on the Kolmogorov distance in [32]; see also [13, 14, 36].

Beyond random graph theory, subgraph counting serves as a powerful tool in network science, enabling the quantification of specific subgraphs (e.g. triangles, cycles, cliques, and trees) to uncover the structural and functional properties of complex systems (see, e.g., [23]). Applications range from identifying subgraphs in biological networks [2] to analyzing the dynamics of functional brain networks [7], providing insights into both local and global organization in real-world networks. However, subgraph counting is computationally intensive, particularly for large networks (see, e.g., [39]). For a comprehensive overview of subgraph counting methods and efficient algorithms in network science, we refer to [33].

Received 8 April 2025; accepted 1 March 2026.
* Postal address: Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China.
** Email address: fengqq@ustc.edu.cn

© The Author(s), 2026. Published by Cambridge University Press on behalf of Applied Probability Trust. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press

Although the ER random graph model has played a pivotal role in the development of random graph theory and network science, its simplicity and restrictive assumptions limit its applicability to modeling real-world networks. Real-world networks exhibit complex topological features that are not captured by the ER random graph model [26]. One such feature is degree heterogeneity [15], which refers to the variability in vertex degrees within a network. Unlike ER random graphs, where vertex degrees are concentrated around the average degree, many real-world networks exhibit heavy-tailed degree distributions, often approximated by power-law behavior (see, e.g., [3, 12]). This indicates the presence of a few highly connected hubs and many sparsely connected vertices. Another key feature is the clustering coefficient, which quantifies the tendency of nodes to form tightly knit groups [27]. In real-world networks, the clustering coefficient is typically high and remains relatively constant as the network grows, whereas it is notably low in ER graphs with comparable edge density (see, e.g., [25]). In other words, real-world networks are usually sparse but contain a large number of triangles. For example, in social networks, the probability that two friends of a given individual are also friends with each other is significantly higher than predicted by the ER random graph model [41].

The β-model is a widely studied statistical network model that incorporates vertex-specific parameters to capture degree heterogeneity in real-world networks [10]. It belongs to the broader class of exponential random graph models, which describe the probability of observing a given network structure (see, e.g., [17, 35]). For statistical inference on the parameters of the β-model and its variants, we refer to [9, 11, 20, 34, 42].

To the best of our knowledge, in contrast to ER random graphs with the regularity of vertex degrees that could facilitate analysis, there are only a few theoretical results on subgraph counting in heterogeneous network models. Several limit laws of the clustering coefficient, which is closely related to the number of triangles, in the configuration model and rank-1 inhomogeneous random graph model for scale-free networks have been established [40, Theorems 1.3–1.5]. Employing the generalized U-statistics, the number of fixed-size cliques (where triangles are a special case) in graphon-based random graphs has also been considered in [16]; see [5] for general subgraph counts. Owing to the inhomogeneous structure of these models, these limit theorems do not yield explicit rates of convergence. In this paper, we aim to derive the asymptotic properties of triangle counts in the sparse β-model as the number of vertices tends to infinity. In particular, using the Malliavin–Stein method [28], under several mild conditions we establish the asymptotic normality of triangle counts with an explicit Berry–Esseen bound.

Throughout this paper, we shall use the following notation. Let $[n]$ denote the set $\{1, 2, \ldots, n\}$ for any positive integer n. We use $C > 0$ to denote a generic constant, which may vary from one occasion to another. For a finite set S, let $|S|$ be the cardinality of S. For a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, denote by $\|\boldsymbol{x}\|_s = (|x_1|^s + |x_2|^s + \cdots + |x_n|^s)^{1/s}$ the L_s-norm of $\boldsymbol{x}$ for $s > 0$, and by $x_{\max}$ and $x_{\min}$ the maximal and minimal entry of $\boldsymbol{x}$, respectively.

The rest of this paper is organized as follows. In Section 2, we first formally introduce the β-model, and then state our main result demonstrating the asymptotic normality of triangle counts under mild sparsity and heterogeneity conditions. In Section 3, we derive the asymptotic mean and variance of the triangle count under the sparsity condition. Finally, with an adaption of the Malliavin–Stein method we prove our main results in Section 4.

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press

2. Model description and main results

The β-model is formally defined as follows [10]. Consider a random graph with vertex set $[n]$, where $n \geq 2$ is an integer. The presence of an edge between vertices i and j is determined independently with probability

$$p_{ij} = \frac{e^{\beta_i+\beta_j}}{1+e^{\beta_i+\beta_j}}, \quad 1 \leq i \neq j \leq n, \tag{1}$$

where $\beta_i \in \mathbb{R}$ is the degree heterogeneity parameter for vertex i. We note that β_i can be negative and may depend on n. These parameters quantify the connectivity propensity of vertices: a larger β_i corresponds to a higher likelihood of vertex i forming edges. Notably, when $\beta_i = \beta$ is a constant for all $i \in [n]$, the β-model reduces to the ER random graph model with edge probability $p = e^{2\beta}/(1+e^{2\beta})$.

Let I_{ij} denote the indicator for the presence of an edge between vertices i and j. By construction, $I_{ii} = 0$, $I_{ij} = I_{ji}$, and the collection $\{I_{ij}, 1 \leq i < j \leq n\}$ consists of independent Bernoulli random variables with success rates p_{ij}. Clearly, the adjacency matrix $\boldsymbol{A} = (I_{ij})_{n\times n}$ is symmetric with zero diagonal. In terms of $\boldsymbol{A}$, the number of triangles T_n in the β-model on n vertices is given by

$$T_n = \frac{1}{6}\mathrm{tr}(\boldsymbol{A}^3) = \sum_{1\leq i<j<k\leq n} I_{ij}I_{jk}I_{ki}, \tag{2}$$

where $\mathrm{tr}(\cdot)$ denotes the matrix trace.

For simplicity of notation, we define $\mu_i = e^{\beta_i} > 0$ for each $i \in [n]$ and let

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^\top,$$

omitting the subscript n. The edge probabilities in (1) then become

$$p_{ij} = \frac{\mu_i\mu_j}{1+\mu_i\mu_j}. \tag{3}$$

Deriving asymptotic properties of T_n under general parameters β_i presents significant challenges due to degree heterogeneity. To address this, we impose the following sparsity conditions:

$$\mu_{\max} \to 0 \quad \text{and} \quad \|\boldsymbol{\mu}\|_2 \to \infty. \tag{4}$$

The first condition ensures network sparsity, while the second guarantees a non-degenerate graph structure with a substantial number of triangles (see Proposition 1). Note that, for any $s > t > 0$, $\|\boldsymbol{\mu}\|_s^s \leq \mu_{\max}^{s-t}\|\boldsymbol{\mu}\|_t^t$, since each μ_i is positive. It thus follows by the assumption $\mu_{\max} \to 0$ in (4) that

$$\|\boldsymbol{\mu}\|_s^s = o\big(\|\boldsymbol{\mu}\|_t^t\big), \quad s > t > 0. \tag{5}$$

The Kolmogorov distance between random variables X and Y is defined as

$$d_K(X, Y) = \sup_{z\in\mathbb{R}} |\mathbb{P}(X \leq z) - \mathbb{P}(Y \leq z)|.$$

Define the normalized triangle count F_n as

$$F_n = \frac{T_n - \mathbb{E}[T_n]}{\sqrt{\mathrm{Var}[T_n]}}. \tag{6}$$

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press

The following theorem provides an upper bound for $d_K(F_n, \mathcal{N})$, where $\mathcal{N}$ denotes a standard normal random variable.

Theorem 1. *Under the sparsity condition* (4) *we have*

$$d_K(F_n, \mathcal{N}) \le \frac{C}{\|\boldsymbol{\mu}\|_2^{5/2}\big(\|\boldsymbol{\mu}\|_3^6 + \|\boldsymbol{\mu}\|_2^2\big)} \sum_{\ell=1}^{5} A_\ell,$$

where $C > 0$ *is a constant, and*

$$\begin{aligned} &A_1 = \|\boldsymbol{\mu}\|_{3/2}^{3/4}\|\boldsymbol{\mu}\|_{5/2}^{5/4}, \qquad A_2 = \|\boldsymbol{\mu}\|_{3/2}^{3/4}\|\boldsymbol{\mu}\|_2^{1/2}\|\boldsymbol{\mu}\|_{7/2}^{7/4}\|\boldsymbol{\mu}\|_4^2, \\ &A_3 = \|\boldsymbol{\mu}\|_2^{5/2}\|\boldsymbol{\mu}\|_5^5, \qquad A_4 = \|\boldsymbol{\mu}\|_2^{3/2}\|\boldsymbol{\mu}\|_{5/2}^{5/2}\|\boldsymbol{\mu}\|_5^{5/2}, \qquad A_5 = \|\boldsymbol{\mu}\|_{7/4}^{7/4}\|\boldsymbol{\mu}\|_{7/2}^{7/4}. \end{aligned} \tag{7}$$

The bound in Theorem 1 is complicated due to the generality of $\boldsymbol{\mu}$. The following example demonstrates that, in specific cases, this bound matches the performance of existing results.

Example 1. Fix a positive integer K and a sequence of positive constants $\{\pi_r, 1 \le r \le K\}$ satisfying $\sum_{r=1}^{K} \pi_r = 1$. Let $\{V_r, 1 \le r \le K\}$ be a partition of [n] with the cardinality $|V_r| = \pi_r n + o(n)$. Assume that, for each $i \in V_r$ with $1 \le r \le K$, $\mu_i = \theta_r n^{-\alpha/2}$, where the constants $\theta_r > 0$ and $0 < \alpha < 1$. Then $\mu_{\max}$ is of order $n^{-\alpha/2}$ and, for any fixed $s > 0$,

$$\|\boldsymbol{\mu}\|_s^s = \sum_{r=1}^{K} |V_r|\, \theta_r^s n^{-\alpha s/2} = n^{1-\alpha s/2}\left(\sum_{r=1}^{K} \pi_r \theta_r^s\right)(1 + o(1)).$$

In particular, when $s = 2$, the quantity $\|\boldsymbol{\mu}\|_2^2$ is of order $n^{1-\alpha}$.

This blockwise-constant scaling is standard in sparse stochastic block models and their degree-corrected variants; see, e.g., [1, 17, 19, 24] for more background. An application of Theorem 1 together with some basic calculations yields $d_K(F_n, \mathcal{N}) \le C\, n^{-\eta(\alpha)}$, where

$$\eta(\alpha) = \begin{cases} 1 - \alpha, & 0 < \alpha \le \frac{1}{2}, \\ \frac{3}{4} - \frac{1}{2}\alpha, & \frac{1}{2} < \alpha \le \frac{2}{3}, \\ \frac{5}{4}(1 - \alpha), & \frac{2}{3} < \alpha < 1. \end{cases}$$

In the special case $K = 1$, we have $\mu_i \equiv \mu = cn^{-\alpha/2}$ for all $i \in [n]$, where $c > 0$ and $\alpha \in (0, 1)$. Then our model reduces to the ER random graph model with edge probability

$$p = \frac{\mu^2}{1 + \mu^2} \sim c^2 n^{-\alpha}$$

as $n \to \infty$. In this case, the order $\eta(\alpha)$ coincides with that established in [22, Theorem 1.1].

Notably, the upper bound in Theorem 1 may not vanish as $n \to \infty$ under the sparsity condition (4), primarily due to the $L_{3/2}$-norm of the $\boldsymbol{\mu}$ term in A_1 and A_2, which can grow very fast. To establish asymptotic normality for T_n, we introduce two additional heterogeneity conditions:

$$\|\boldsymbol{\mu}\|_{3/2}^{3/2} = O\big(\|\boldsymbol{\mu}\|_2^6\big), \tag{8}$$

$$\frac{\mu_{\max}}{\mu_{\min}} = O\big(\|\boldsymbol{\mu}\|_2^{3/2}\big). \tag{9}$$

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press

Condition (8) directly restricts the growth rate of the $L_{3/2}$-norm of $\boldsymbol{\mu}$. Condition (9) accommodates considerable degree heterogeneity in the β-model, thereby permitting the ratio $\mu_{\max}/\mu_{\min}$ to diverge at an appropriate rate. This is in sharp contrast to the ER random graph model, where this ratio is invariably 1. Such regularity conditions are widely adopted in statistical network analysis. For instance, in community detection problems, several constraints on the extreme values of degree heterogeneity parameters are imposed to bound the estimation error (see, e.g., [18]).

The following result establishes the asymptotic normality of T_n under these conditions.

Theorem 2. *Under the sparsity condition* (4)*, if either* (8) *or* (9) *holds, then* $F_n \xrightarrow{\mathrm{D}} \mathcal{N}$ *as* $n \to \infty$*, where* $\xrightarrow{\mathrm{D}}$ *denotes convergence in distribution.*

3. Mean and variance

In this section, we derive the asymptotic mean and variance of T_n, the number of triangles in the β-model, under the sparsity condition (4).

Proposition 1. *Suppose that the sparsity condition* (4) *holds. As* $n \to \infty$*, we have*

$$\mathbb{E}[T_n] = \frac{1}{6}\|\boldsymbol{\mu}\|_2^6(1+o(1)), \tag{10}$$

$$\mathrm{Var}[T_n] = \frac{1}{6}\|\boldsymbol{\mu}\|_2^4\big(3\|\boldsymbol{\mu}\|_3^6 + \|\boldsymbol{\mu}\|_2^2\big)(1+o(1)). \tag{11}$$

Proof. We first note that, by (3) and (4), the edge probability p_{ij} satisfies

$$p_{ij} = (1+o(1))\mu_i\mu_j, \quad 1 \le i \ne j \le n. \tag{12}$$

For distinct vertices $i, j, k \in [n]$, by the definition of the β-model, the indicators I_{ij}, I_{jk}, and I_{ki} are independent Bernoulli random variables. Using (2) and (12), the expectation of T_n is equal to

$$\mathbb{E}[T_n] = \sum_{1\le i<j<k\le n} p_{ij}p_{jk}p_{ki} = (1+o(1)) \sum_{1\le i<j<k\le n} \mu_i^2\mu_j^2\mu_k^2. \tag{13}$$

In terms of the norms of $\boldsymbol{\mu}$, the summation on the right-hand side of (13) can be rewritten as

$$\begin{aligned}\sum_{1\le i<j<k\le n} \mu_i^2\mu_j^2\mu_k^2 &= \frac{1}{6}\sum_{1\le i,j,k\le n} \mu_i^2\mu_j^2\mu_k^2 - \frac{1}{2}\sum_{1\le i,j\le n} \mu_i^2\mu_j^4 + \frac{1}{3}\sum_{i=1}^n \mu_i^6 \\ &= \frac{1}{6}\|\boldsymbol{\mu}\|_2^6 - \frac{1}{2}\|\boldsymbol{\mu}\|_2^2\|\boldsymbol{\mu}\|_4^4 + \frac{1}{3}\|\boldsymbol{\mu}\|_6^6.\end{aligned}$$

Combining this with (5) and (13) yields (10).

We now turn to the variance of T_n. To this end, we define $X_{ij} = I_{ij} - p_{ij}$, $1 \le i \ne j \le n$. Then, for any $1 \le i \ne j \le n$, we have $\mathbb{E}[X_{ij}] = 0$, and by (12),

$$\mathrm{Var}[X_{ij}] = \mathbb{E}\big[X_{ij}^2\big] = p_{ij}(1-p_{ij}) = (1+o(1))\mu_i\mu_j. \tag{14}$$

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press

By (13), we can now reformulate (2) as

$$\begin{aligned} T_n &= \sum_{1\le i<j<k\le n} (X_{ij}+p_{ij})(X_{jk}+p_{jk})(X_{ki}+p_{ki}) \\ &= \mathbb{E}[T_n] + \sum_{1\le i<j\le n} c_{ij}X_{ij} + \sum_{\substack{1\le i<j\le n \\ k\ne i,j}} p_{ij}X_{jk}X_{ki} + \sum_{1\le i<j<k\le n} X_{ij}X_{jk}X_{ki}, \end{aligned} \tag{15}$$

where, by (12),

$$c_{ij} = \sum_{k\ne i,j} p_{jk}p_{ki} = (1+o(1))\mu_i\mu_j \sum_{k\ne i,j} \mu_k^2 = (1+o(1))\|\boldsymbol{\mu}\|_2^2\mu_i\mu_j.$$

The $1+\binom{n}{2}+(n-2)\binom{n}{2}+\binom{n}{3}$ terms on the right-hand side of (15) are pairwise uncorrelated. By (12) and (14), we have

$$\begin{aligned} \mathrm{Var}[T_n] &= \sum_{1\le i\le j\le n} c_{ij}^2\mathbb{E}[X_{ij}^2] + \sum_{\substack{1\le i<j\le n \\ k\ne i,j}} p_{ij}^2\mathbb{E}[X_{jk}^2]\mathbb{E}[X_{ki}^2] + \sum_{1\le i<j<k\le n} \mathbb{E}[X_{ij}^2]\mathbb{E}[X_{jk}^2]\mathbb{E}[X_{ki}^2] \\ &= (1+o(1))\left(\|\boldsymbol{\mu}\|_2^4 \sum_{1\le i<j\le n} \mu_i^3\mu_j^3 + \|\boldsymbol{\mu}\|_2^2 \sum_{1\le i<j\le n} \mu_i^3\mu_j^3 + \sum_{1\le i<j<k\le n} \mu_i^2\mu_j^2\mu_k^2\right). \end{aligned}$$

Since $\|\boldsymbol{\mu}\|_2^2 = o(\|\boldsymbol{\mu}\|_2^4)$ and $\sum_{1\le i<j\le n}\mu_i^3\mu_j^3 = \frac{1}{2}\big(\|\boldsymbol{\mu}\|_3^6 - \|\boldsymbol{\mu}\|_6^6\big)$, by (10) and (13) we can proceed with

$$\begin{aligned} \mathrm{Var}[T_n] &= \left(\frac{1}{2}\|\boldsymbol{\mu}\|_2^4\big(\|\boldsymbol{\mu}\|_3^6 - \|\boldsymbol{\mu}\|_6^6\big) + \frac{1}{6}\|\boldsymbol{\mu}\|_2^6\right)(1+o(1)) \\ &= \frac{1}{6}\|\boldsymbol{\mu}\|_2^4\big(3\|\boldsymbol{\mu}\|_3^6 - 3\|\boldsymbol{\mu}\|_6^6 + \|\boldsymbol{\mu}\|_2^2\big)(1+o(1)), \end{aligned}$$

which, together with (5), completes the proof of (11). □

An immediate consequence of Proposition 1 is the following.

Proposition 2. *Under the sparsity condition* (4)*, as* $n\to\infty$*,* $T_n/\|\boldsymbol{\mu}\|_2^6 \xrightarrow{\mathrm{P}} \frac{1}{6}$*, where* $\xrightarrow{\mathrm{P}}$ *denotes convergence in probability.*

Proof. Applying Proposition 1 and Chebyshev's inequality, for any $\varepsilon>0$, by (5) we have

$$\mathbb{P}\left(\left|\frac{T_n}{\mathbb{E}[T_n]} - 1\right| > \varepsilon\right) \le \frac{\mathrm{Var}[T_n]}{(\varepsilon\mathbb{E}[T_n])^2} = \frac{6\big(3\|\boldsymbol{\mu}\|_3^6 + \|\boldsymbol{\mu}\|_2^2\big)}{\varepsilon^2\|\boldsymbol{\mu}\|_2^8}(1+o(1)) \to 0,$$

which implies that $T_n/\mathbb{E}[T_n]$ converges to 1 in probability. Thus, the desired result follows via Slutsky's theorem and (10). □

4. Proofs

This section is devoted to the proofs of our main results stated in Section 2. We first introduce several auxiliary lemmas essential to our approach.

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press

Our primary tool for establishing the asymptotic normality of triangle counts in the β-model is the Malliavin–Stein method [28], a powerful synthesis of Malliavin calculus [30] and Stein's method (see, e.g., [4]). This method is particularly effective for analyzing normal approximations when applied to independent Rademacher or Bernoulli random variables [21, 29]. In particular, by applying this method Berry–Esseen bounds for triangle counts in ER random graphs have been established [22], while non-uniform Berry–Esseen bounds for counts of any fixed subgraph have been further derived in [8]. For more background and related bounds in the Malliavin–Stein framework, see, e.g., [14] and references therein.

To state the normal approximation bound used in our approach (Lemma 1), we introduce, following [14], the discrete gradient operator for functionals of independent Bernoulli random variables. For our purposes, only the first- and second-order discrete gradients are needed.

Let $m \geq 3$ be an integer, and $\boldsymbol{X} = (X_1, \ldots, X_m)$ a random vector of independent Bernoulli random variables with $\mathbb{P}(X_a = 1) = p_a$ and $\mathbb{P}(X_a = 0) = q_a$, $a \in [m]$, where $0 < p_a < 1$ and $q_a = 1 - p_a$. For a measurable function $f\colon \{0, 1\}^m \to \mathbb{R}$ and $F = f(\boldsymbol{X})$, define the discrete gradient of F with respect to X_a by

$$D_a F = \sqrt{p_a q_a}\big(f(\boldsymbol{X}_a^+) - f(\boldsymbol{X}_a^-)\big), \tag{16}$$

where

$$\boldsymbol{X}_a^+ = (X_1, \ldots, X_{a-1}, 1, X_{a+1}, \ldots, X_m), \qquad \boldsymbol{X}_a^- = (X_1, \ldots, X_{a-1}, 0, X_{a+1}, \ldots, X_m).$$

Using (16), we can further define the second-order discrete gradient $D_b D_a F := D_b(D_a F)$. Note that for any $a, b \in [m]$, $D_b D_a F = D_a D_b F$.

The following lemma plays an important role in our proof of Theorem 1.

Lemma 1. *Let $F = f(\boldsymbol{X})$ be a random variable with $\mathbb{E}[F] = 0$ and $\mathrm{Var}[F] = 1$. Then*

$$d_K(F, \mathcal{N}) \leq C \sum_{k=1}^{5} \sqrt{B_k},$$

where $\mathcal{N}$ is a standard normal random variable, and

$$B_1 = \sum_{a,b,c \in [m]} \sqrt{\mathbb{E}\big[D_a^2 D_b^2\big]\mathbb{E}\big[D_{ca}^2 D_{cb}^2\big]},$$

$$B_2 = \sum_{a,b,c \in [m]} \frac{1}{p_c q_c}\mathbb{E}\big[D_{ca}^2 D_{cb}^2\big], \qquad B_3 = \sum_{a=1}^{m} \frac{1}{p_a q_a}\mathbb{E}\big[D_a^4\big],$$

$$B_4 = \sum_{a,b \in [m]} \frac{1}{p_a q_a}\sqrt{\mathbb{E}\big[D_a^4\big]\mathbb{E}\big[D_{ab}^4\big]}, \qquad B_5 = \sum_{a,b \in [m]} \frac{1}{p_a q_a p_b q_b}\mathbb{E}\big[D_{ab}^4\big],$$

with $D_a := D_a F$ and $D_{ab} := D_a D_b F$.

Proof. This lemma follows directly by applying [14, Theorem 4.1(i)] to the independent Rademacher random variables $\{Y_a = 2X_a - 1\colon a \in [m]\}$. We only need to verify [14, Condition (3.3)], and that $u\colon a \mapsto (p_a q_a)^{-1/2} D_a F\, |D_a L^{-1} F|$ belongs to $\mathrm{Dom}(\delta)$, where $\mathrm{Dom}(\delta)$ and the operator L are defined as in [14, Section 2].

Note that u depends only on finitely many independent Rademacher random variables $\{Y_a, a \in [m]\}$. [14, Condition (3.3)] is satisfied by [14, Remark 3.2(ii)]. Moreover, by the

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press

Wiener–Itô–Walsh decomposition, the chaos expansion of u contains only finitely many terms (see, for instance, [31, 444]). Consequently, [22, Condition (2.14)] is satisfied, which implies that $u \in \mathrm{Dom}(\delta)$. □

To apply Lemma 1, we denote the set of all possible edges in the β-model on n vertices by $\{e_1, e_2, \ldots, e_m\}$, where $m = \binom{n}{2}$. For each edge e_a with $a \in [m]$, let $1 \le i_a < j_a \le n$ denote its endpoints. Note that the edge indicators $\{I_{i_aj_a}, a \in [m]\}$ are independent Bernoulli random variables with $\mathbb{P}(I_{i_aj_a} = 1) = p_{i_aj_a}$ and $\mathbb{P}(I_{i_aj_a} = 0) = 1 - p_{i_aj_a}$. Hence, the normalized triangle count F_n in (6) is a measurable function of these variables and is thus amenable to analysis via Lemma 1.

By (16), we have

$$D_a := D_a F_n = \frac{\sqrt{h_a}}{\sqrt{\mathrm{Var}[T_n]}} V_a, \quad a \in [m], \tag{17}$$

where

$$h_a = p_{i_aj_a}(1 - p_{i_aj_a}), \tag{18}$$

$$V_a = \sum_{k \neq i_a, j_a} I_{i_ak} I_{j_ak}. \tag{19}$$

In other words, V_a counts the number of wedges (i.e. paths of length two) with endpoints i_a and j_a. Graphically, the gradient D_aF_n measures the sensitivity of triangle counts to flipping the status of e_a.

For the second-order gradient, consider any two given edges $e_a = \{i_a, j_a\}$ and $e_b = \{i_b, j_b\}$. By definition, we can immediately obtain that $D_{ab} = 0$ if either e_a and e_b are an identical edge or disjoint (i.e. the intersection $\{i_a, j_a\} \cap \{i_b, j_b\}$ is the empty set $\varnothing$). If $|\{i_a, j_a\} \cap \{i_b, j_b\}| = 1$, then (16) implies that

$$D_b V_a = \sqrt{h_b}\, I_{\ell_1 \ell_2}, \tag{20}$$

where $\{\ell_1, \ell_2\} = (\{i_a, j_a\} \cup \{i_b, j_b\}) \setminus (\{i_a, j_a\} \cap \{i_b, j_b\})$, i.e., ℓ_1 and ℓ_2 are the endpoints of e_a and e_b excluding their common vertex. Consequently, it follows by (17) and (20) that, for any $a, b \in [m]$,

$$D_{ab} := D_a D_b F_n = D_b D_a F_n = \frac{\sqrt{h_a h_b}}{\sqrt{\mathrm{Var}[T_n]}} \Delta_{ab}, \tag{21}$$

with

$$\Delta_{ab} = \begin{cases} 0 & \text{if } \{i_a, j_a\} = \{i_b, j_b\} \text{ or } \{i_a, j_a\} \cap \{i_b, j_b\} = \varnothing, \\ I_{\ell_1 \ell_2} & \text{if } |\{i_a, j_a\} \cap \{i_b, j_b\}| = 1. \end{cases} \tag{22}$$

In addition to Lemma 1, we further require the following auxiliary lemmas.

Lemma 2. *For any $s \ge 2$, $\mathbb{E}[V_a^s] \le C\big(\|\boldsymbol{\mu}\|_2^2 \mu_{i_a}\mu_{j_a} + \|\boldsymbol{\mu}\|_2^{2s} \mu_{i_a}^s \mu_{j_a}^s\big)$, where $C = C(s) > 0$ is a constant depending only on s.*

Proof. We begin with the inequality $\mathbb{E}[|X + Y|^s] \le 2^{s-1}\mathbb{E}[|X|^s + |Y|^s]$, which holds for any random variables X and Y. Applying this to V_a given in (19), we have

$$\mathbb{E}[V_a^s] \le C\mathbb{E}\left[\left|\sum_{k \neq i_a, j_a} (I_{i_ak} I_{j_ak} - p_{i_ak} p_{j_ak})\right|^s + \left(\sum_{k \neq i_a, j_a} p_{i_ak} p_{j_ak}\right)^s\right]. \tag{23}$$

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press

To bound the first term on the right-hand side of (23), observe that $I_{i_ak}I_{j_ak}$ are independent Bernoulli variables with success rate $p_{i_ak}p_{j_ak}$ for $k \neq i_a, j_a$. By Rosenthal's inequality we have

$$\begin{aligned}\mathbb{E}\left|\sum_{k\neq i_a,j_a}(I_{i_ak}I_{j_ak}-p_{i_ak}p_{j_ak})\right|^s &\le C\Bigg[\sum_{k\neq i_a,j_a}\mathbb{E}|I_{i_ak}I_{j_ak}-p_{i_ak}p_{j_ak}|^s \\ &\quad+\left(\sum_{k\neq i_a,j_a}\mathrm{Var}[I_{i_ak}I_{j_ak}]\right)^{s/2}\Bigg] \\ &\le C\left[\sum_{k\neq i_a,j_a}p_{i_ak}p_{j_ak}+\left(\sum_{k\neq i_a,j_a}p_{i_ak}p_{j_ak}\right)^{s/2}\right].\end{aligned}$$

Substituting this into (23) and using the inequality $x^{s/2} \le x + x^s$ for $x > 0$ and $s \ge 2$, by (12) we thus have

$$\begin{aligned}\mathbb{E}[V_a^s] &\le C\left[\sum_{k\neq i_a,j_a}p_{i_ak}p_{j_ak}+\left(\sum_{k\neq i_a,j_a}p_{i_ak}p_{j_ak}\right)^s\right] \\ &\le C\left[\mu_{i_a}\mu_{j_a}\sum_{k\neq i_a,j_a}\mu_k^2+\left(\mu_{i_a}\mu_{j_a}\sum_{k\neq i_a,j_a}\mu_k^2\right)^s\right] \le C\big(\|\boldsymbol{\mu}\|_2^2\mu_{i_a}\mu_{j_a}+\|\boldsymbol{\mu}\|_2^{2s}\mu_{i_a}^s\mu_{j_a}^s\big),\end{aligned}$$

which completes the proof of Lemma 2. □

The following inequality is used repeatedly in our analysis.

Lemma 3. *Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ be a positive vector (i.e. $x_i > 0$ for all $i \in [n]$). For any $s, t > 0$, $\|\boldsymbol{x}\|_{(s+t)/2}^{s+t} \le \|\boldsymbol{x}\|_s^s \|\boldsymbol{x}\|_t^t$.*

Proof. The inequality follows directly from the Cauchy–Schwarz inequality. □

With the above preparation, we proceed to prove Theorem 1.

Proof of Theorem 1. Recall the second-order discrete gradients D_{ab} given in (21) for all $a, b \in [m]$, where $m = \binom{n}{2}$ represents the number of all possible edges in the β-model on n vertices. Observe that $\Delta_{ab}^s = \Delta_{ab} = \Delta_{ba}$ for all $s > 0$, and Δ_{ca} and Δ_{cb} are independent for distinct edges e_a, e_b, e_c. Combining Lemma 1 with (17) and (21), we obtain

$$d_K(F_n, \mathcal{N}) \le \frac{C}{\mathrm{Var}[T_n]}\sum_{k=1}^5\sqrt{\widetilde{B}_k}, \tag{24}$$

where

$$\widetilde{B}_1 := \sum_{a,b,c\in[m]} h_a h_b h_c \sqrt{\mathbb{E}\big[V_a^2V_b^2\big]\mathbb{E}[\Delta_{ca}\Delta_{cb}]},$$

$$\widetilde{B}_2 := \sum_{a,b,c\in[m]} h_a h_b h_c \mathbb{E}[\Delta_{ca}\Delta_{cb}], \qquad \widetilde{B}_3 := \sum_{a=1}^m h_a \mathbb{E}\big[V_a^4\big],$$

$$\widetilde{B}_4 := \sum_{a,b\in[m]} h_a h_b \sqrt{\mathbb{E}\big[V_a^4\big]\mathbb{E}[\Delta_{ab}]}, \qquad \widetilde{B}_5 := \sum_{a,b\in[m]} h_a h_b \mathbb{E}[\Delta_{ab}].$$

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press

To prove Theorem 1, by (11) and (24) it suffices to show that $\sum_{k=1}^{5}\sqrt{\widetilde{B}_k} \le C\|\boldsymbol{\mu}\|_2^{3/2}\sum_{\ell=1}^{5}A_\ell$, where $A_\ell > 0$ ($\ell = 1, 2, \ldots, 5$) are defined in (7). Applying the Cauchy–Schwarz inequality, a sufficient condition for this to hold is

$$\sum_{k=1}^{5}\widetilde{B}_k \le C\|\boldsymbol{\mu}\|_2^3\sum_{\ell=1}^{5}A_\ell^2. \tag{25}$$

Thus, the remainder of the proof is devoted to verifying (25).

We now analyze the terms $\widetilde{B}_k$ individually, beginning with $\widetilde{B}_3$. From (12) and (18), we deduce that $h_a = (1+o(1))\mu_{i_a}\mu_{j_a}$, $a \in [m]$. Applying Lemma 2 with $s = 4$ yields

$$\widetilde{B}_3 \le C\sum_{a=1}^{m}\left(\|\boldsymbol{\mu}\|_2^2\mu_{i_a}^2\mu_{j_a}^2 + \|\boldsymbol{\mu}\|_2^8\mu_{i_a}^5\mu_{j_a}^5\right).$$

Since, for any $t > 0$,

$$\sum_{a=1}^{m}\mu_{i_a}^t\mu_{j_a}^t = \sum_{1\le i<j\le n}\mu_i^t\mu_j^t = \frac{1}{2}\left[\left(\sum_{i=1}^{n}\mu_i^t\right)^2 - \sum_{i=1}^{n}\mu_i^{2t}\right] \le \frac{1}{2}\|\boldsymbol{\mu}\|_t^{2t},$$

we have

$$\widetilde{B}_3 \le C\left(\|\boldsymbol{\mu}\|_2^6 + \|\boldsymbol{\mu}\|_2^8\|\boldsymbol{\mu}\|_5^{10}\right) = C\|\boldsymbol{\mu}\|_2^3\left(\|\boldsymbol{\mu}\|_2^3 + A_3^2\right). \tag{26}$$

Next, consider $\widetilde{B}_5$, which involves pairs of edges $e_a = (i_a, j_a)$ and $e_b = (i_a, j_b)$ sharing a common vertex i_a. By (12) and (22), we have

$$\begin{aligned}
\widetilde{B}_5 &= \sum_{i_a=1}^{n}\sum_{j_a\ne i_a}\sum_{j_b\ne i_a, j_a} p_{i_aj_b}(1-p_{i_aj_a})p_{i_aj_b}(1-p_{i_aj_b})p_{j_aj_b} \\
&= \sum_{i=1}^{n}\sum_{j\ne i}\sum_{k\ne i,j} p_{ij}(1-p_{ij})p_{ik}(1-p_{ik})p_{jk} \\
&\le \sum_{i=1}^{n}\sum_{j\ne i}\sum_{k\ne i,j} p_{ij}p_{ik}p_{jk} \le \sum_{i,j,k\in[n]}\mu_i^2\mu_j^2\mu_k^2 = \|\boldsymbol{\mu}\|_2^6,
\end{aligned} \tag{27}$$

where in the second equality we relabeled the indices i_a, j_a, j_b as i, j, k for simplicity, and in the last inequality we applied the bound $p_{ij} \le \mu_i\mu_j$.

Invoking Lemma 3 with $s = \frac{5}{2}$ and $t = \frac{3}{2}$ gives $\|\boldsymbol{\mu}\|_2^4 \le \|\boldsymbol{\mu}\|_{3/2}^{3/2}\|\boldsymbol{\mu}\|_{5/2}^{5/2} = A_1^2$. Combining this with (26) and (27), and noting that $\|\boldsymbol{\mu}\|_2^3 = o(\|\boldsymbol{\mu}\|_2^4)$ under our assumptions, we have

$$\widetilde{B}_3 + \widetilde{B}_5 \le C\|\boldsymbol{\mu}\|_2^3\left(A_1^2 + A_3^2\right). \tag{28}$$

For $\widetilde{B}_4$, Lemma 2 with $s = 4$ provides

$$\sqrt{\mathbb{E}[V_a^4]} \le C\left(\|\boldsymbol{\mu}\|_2\mu_{i_a}^{1/2}\mu_{j_a}^{1/2} + \|\boldsymbol{\mu}\|_2^4\mu_{i_a}^2\mu_{j_a}^2\right). \tag{29}$$

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press

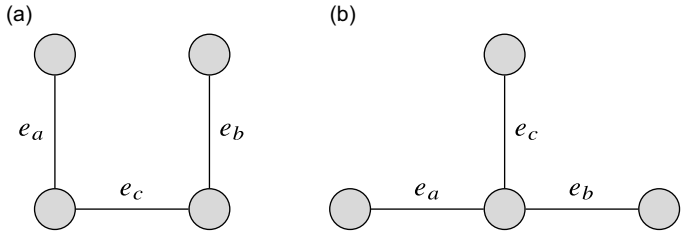


FIGURE 1. Edge e_c shares a common vertex with two other edges, e_a and e_b.

Following the approach for $\widetilde{B}_5$, by (29) we bound $\widetilde{B}_4$ as

$$\begin{aligned}\widetilde{B}_4 &\le C\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j} p_{ij}p_{ik}\Big(\|\boldsymbol{\mu}\|_2\mu_i^{1/2}\mu_j^{1/2}+\|\boldsymbol{\mu}\|_2^4\mu_i^2\mu_j^2\Big)p_{jk}^{1/2}\\ &\le C\sum_{i,j,k\in[n]}\Big(\|\boldsymbol{\mu}\|_2\mu_i^{5/2}\mu_j^2\mu_k^{3/2}+\|\boldsymbol{\mu}\|_2^4\mu_i^4\mu_j^{7/2}\mu_k^{3/2}\Big)=C\|\boldsymbol{\mu}\|_2^3\big(A_1^2+A_2^2\big),\end{aligned}\tag{30}$$

which, together with (28), implies that

$$\widetilde{B}_3+\widetilde{B}_4+\widetilde{B}_5\le C\|\boldsymbol{\mu}\|_2^3\big(A_1^2+A_2^2+A_3^2\big).\tag{31}$$

We now turn to $\widetilde{B}_2$. By distinguishing cases where edges e_a and e_b are identical or distinct, we decompose this term as

$$\widetilde{B}_2=\sum_{a,c\in[m]}h_a^2h_c\mathbb{E}[\Delta_{ca}]+\sum_{\{a,b,c\}\subset[m]}h_ah_bh_c\mathbb{E}[\Delta_{ca}]\mathbb{E}[\Delta_{cb}]=:\widetilde{B}_{21}+\widetilde{B}_{22},\tag{32}$$

where the simple fact $\mathbb{E}[\Delta_{ca}^2]=\mathbb{E}[\Delta_{ca}]$ is used.

For $\widetilde{B}_{21}$, analogous to the analysis of $\widetilde{B}_5$ and $\widetilde{B}_4$, we obtain

$$\begin{aligned}\widetilde{B}_{21}&=\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}p_{ij}^2(1-p_{ij})^2p_{ik}(1-p_{ik})p_{jk}\\ &\le\sum_{i,j,k\in[n]}p_{ij}^2p_{ik}p_{jk}\le\sum_{i,j,k\in[n]}\mu_i^3\mu_j^3\mu_k^2=\|\boldsymbol{\mu}\|_2^2\|\boldsymbol{\mu}\|_3^6.\end{aligned}\tag{33}$$

To bound $\widetilde{B}_{22}$, we only need to consider triples of edges (e_a, e_b, e_c) where e_c shares a common vertex with both e_a and e_b. Two configurations exist: either e_c connects to e_a and e_b via distinct common vertices (Figure 1(a)), or all three edges share a unique common vertex (Figure 1(b)). This yields

$$\begin{aligned}\widetilde{B}_{22}&=\sum_{1\le i\neq j\le n}\sum_{k\neq i,j}\sum_{\ell\neq i,j,k}\big(p_{ij}(1-p_{ij})p_{ik}(1-p_{ik})p_{j\ell}(1-p_{j\ell})p_{i\ell}p_{jk}\\ &\qquad+p_{ij}(1-p_{ij})p_{ik}(1-p_{ik})p_{i\ell}(1-p_{i\ell})p_{jk}p_{j\ell}\big)\\ &\le 2\sum_{i,j,k,\ell\in[n]}p_{ij}p_{ik}p_{j\ell}p_{i\ell}p_{jk}\le 2\sum_{i,j,k,\ell\in[n]}\mu_i^3\mu_j^3\mu_k^2\mu_\ell^2=2\|\boldsymbol{\mu}\|_2^4\|\boldsymbol{\mu}\|_3^6.\end{aligned}\tag{34}$$

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press

Substituting (33) and (34) into (32) and applying the sparsity condition (4), we obtain

$$\widetilde{B}_2 \le C\big(\|\boldsymbol{\mu}\|_2^2\|\boldsymbol{\mu}\|_3^6 + \|\boldsymbol{\mu}\|_2^4\|\boldsymbol{\mu}\|_3^6\big) \le C\|\boldsymbol{\mu}\|_2^4\|\boldsymbol{\mu}\|_3^6. \tag{35}$$

Invoking Lemma 3 with $s=2$ and $t=4$ gives $\|\boldsymbol{\mu}\|_3^6 \le \|\boldsymbol{\mu}\|_2^2\|\boldsymbol{\mu}\|_4^4$. Utilizing the asymptotic relations $\|\boldsymbol{\mu}\|_2 = o\big(\|\boldsymbol{\mu}\|_2^2\big)$, $\|\boldsymbol{\mu}\|_4^4 = o\big(\|\boldsymbol{\mu}\|_{7/2}^{7/2}\big)$, and $\|\boldsymbol{\mu}\|_2^4 = o\big(\|\boldsymbol{\mu}\|_{7/4}^{7/2}\big)$, we have

$$\|\boldsymbol{\mu}\|_2\|\boldsymbol{\mu}\|_3^6 \le \|\boldsymbol{\mu}\|_2^3\|\boldsymbol{\mu}\|_4^4 = o\big(\|\boldsymbol{\mu}\|_2^4\|\boldsymbol{\mu}\|_{7/2}^{7/2}\big) = o\big(A_5^2\big).$$

Thus, from (35) we can conclude that

$$\widetilde{B}_2 = o\big(\|\boldsymbol{\mu}\|_2^3 A_5^2\big). \tag{36}$$

Consider $\widetilde{B}_1$. Following a similar decomposition to (32), we can rewrite it as

$$\begin{aligned}\widetilde{B}_1 &= \sum_{a,c\in[m]} h_a^2 h_c \sqrt{\mathbb{E}\big[V_a^4\big]\mathbb{E}[\Delta_{ca}]} + \sum_{\{a,b,c\}\subset[m]} h_a h_b h_c \sqrt{\mathbb{E}\big[V_a^2 V_b^2\big]\mathbb{E}[\Delta_{ca}]\mathbb{E}[\Delta_{cb}]} \\ &=: \widetilde{B}_{11} + \widetilde{B}_{12}.\end{aligned} \tag{37}$$

Comparing $\widetilde{B}_{11}$ with the expression for $\widetilde{B}_4$, by (30) we obtain

$$\widetilde{B}_{11} = o(\widetilde{B}_4) = o\big(\|\boldsymbol{\mu}\|_2^3\big(A_1^2 + A_2^2\big)\big). \tag{38}$$

For $\widetilde{B}_{12}$, using the Cauchy–Schwarz inequality and (29) gives

$$\begin{aligned}\widetilde{B}_{12} &\le \sum_{\{a,b,c\}\subset[m]} h_a h_b h_c \big(\mathbb{E}\big[V_a^4\big]\big)^{1/4}\big(\mathbb{E}\big[V_b^4\big]\big)^{1/4}\sqrt{\mathbb{E}[\Delta_{ca}]\mathbb{E}[\Delta_{cb}]} \\ &\le C\|\boldsymbol{\mu}\|_2 \sum_{\{a,b,c\}\subset[m]} h_a h_b h_c \Big(\mu_{i_a}^{1/4}\mu_{j_a}^{1/4} + \|\boldsymbol{\mu}\|_2^{3/2}\mu_{i_a}\mu_{j_a}\Big)\Big(\mu_{i_b}^{1/4}\mu_{j_b}^{1/4} + \|\boldsymbol{\mu}\|_2^{3/2}\mu_{i_b}\mu_{j_b}\Big) \\ &\qquad \times \sqrt{\mathbb{E}[\Delta_{ca}]\mathbb{E}[\Delta_{cb}]}.\end{aligned}$$

Then, analogously to (34), we can proceed with

$$\begin{aligned}\widetilde{B}_{12} \le C\|\boldsymbol{\mu}\|_2 \sum_{1\le i\ne j\le n} \sum_{k\ne i,j} \sum_{\ell\ne i,j,k} \Big[& p_{ij}p_{ik}p_{j\ell}\Big(\mu_i^{1/4}\mu_k^{1/4} + \|\boldsymbol{\mu}\|_2^{3/2}\mu_i\mu_k\Big) \\ & \times \Big(\mu_j^{1/4}\mu_\ell^{1/4} + \|\boldsymbol{\mu}\|_2^{3/2}\mu_j\mu_\ell\Big)p_{i\ell}^{1/2}p_{jk}^{1/2} \\ & + p_{ij}p_{ik}p_{i\ell}\Big(\mu_i^{1/4}\mu_k^{1/4} + \|\boldsymbol{\mu}\|_2^{3/2}\mu_i\mu_k\Big) \\ & \times \Big(\mu_i^{1/4}\mu_\ell^{1/4} + \|\boldsymbol{\mu}\|_2^{3/2}\mu_i\mu_\ell\Big)p_{jk}^{1/2}p_{j\ell}^{1/2}\Big]\end{aligned}$$

$$\begin{aligned}&\le C\|\boldsymbol{\mu}\|_2 \\ &\quad \times \sum_{i,j,k,\ell\in[n]} \Big[\mu_i^{5/2}\mu_j^{5/2}\mu_k^{3/2}\mu_\ell^{3/2} \\ &\qquad \times \Big(\mu_i^{1/4}\mu_j^{1/4}\mu_k^{1/4}\mu_\ell^{1/4} + \|\boldsymbol{\mu}\|_2^{3/2}\mu_i\mu_j^{1/4}\mu_k\mu_\ell^{1/4} + \|\boldsymbol{\mu}\|_2^3\mu_i\mu_j\mu_k\mu_\ell\Big) \\ &\qquad + \mu_i^3\mu_j^2\mu_k^{3/2}\mu_\ell^{3/2}\Big(\mu_i^{/12}\mu_k^{/14}\mu_\ell^{/14} + \|\boldsymbol{\mu}\|_2^{3/2}\mu_i^{5/4}\mu_k\mu_\ell^{1/4} + \|\boldsymbol{\mu}\|_2^3\mu_i^2\mu_k\mu_\ell\Big)\Big]\end{aligned}$$

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press

$$= C\|\boldsymbol{\mu}\|_2\Big(\|\boldsymbol{\mu}\|_{7/4}^{7/2}\|\boldsymbol{\mu}\|_{11/4}^{11/2} + \|\boldsymbol{\mu}\|_{7/4}^{7/4}\|\boldsymbol{\mu}\|_2^{3/2}\|\boldsymbol{\mu}\|_{5/2}^{5/2}\|\boldsymbol{\mu}\|_{11/4}^{11/4}\|\boldsymbol{\mu}\|_{7/2}^{7/2} + \|\boldsymbol{\mu}\|_2^3\|\boldsymbol{\mu}\|_{5/2}^5\|\boldsymbol{\mu}\|_{7/2}^7$$
$$+ \|\boldsymbol{\mu}\|_{7/4}^{7/2}\|\boldsymbol{\mu}\|_2^2\|\boldsymbol{\mu}\|_{7/2}^{7/2} + \|\boldsymbol{\mu}\|_{7/4}^{7/4}\|\boldsymbol{\mu}\|_2^{7/2}\|\boldsymbol{\mu}\|_{5/2}^{5/2}\|\boldsymbol{\mu}\|_{17/4}^{17/4} + \|\boldsymbol{\mu}\|_2^5\|\boldsymbol{\mu}\|_{5/2}^5\|\boldsymbol{\mu}\|_5^5\Big), \tag{39}$$

where the second inequality follows from $p_{ij} \le \mu_i\mu_j$ for all $i, j \in [n]$, together with symmetry. We denote the six terms in the final parentheses above in turn by $\widetilde{B}_{12}^{(1)}, \widetilde{B}_{12}^{(2)}, \ldots, \widetilde{B}_{12}^{(6)}$. Clearly, from the relation $\big(\widetilde{B}_{12}^{(2)}\big)^2 = \widetilde{B}_{12}^{(1)}\widetilde{B}_{12}^{(3)}$, we have

$$\widetilde{B}_{12}^{(2)} \le \frac{1}{2}\big(\widetilde{B}_{12}^{(1)} + \widetilde{B}_{12}^{(3)}\big). \tag{40}$$

Setting $s = \frac{7}{2}$ and $t = 5$ in Lemma 3 yields $\|\boldsymbol{\mu}\|_{17/4}^{17/2} \le \|\boldsymbol{\mu}\|_5^5\|\boldsymbol{\mu}\|_{7/2}^{7/2}$, which implies that $\big(\widetilde{B}_{12}^{(5)}\big)^2 \le \widetilde{B}_{12}^{(4)}\widetilde{B}_{12}^{(6)}$. Hence, we also have

$$\widetilde{B}_{12}^{(5)} \le \frac{1}{2}\big(\widetilde{B}_{12}^{(4)} + \widetilde{B}_{12}^{(6)}\big) = \frac{1}{2}\|\boldsymbol{\mu}\|_2^2\big(A_5^2 + A_4^2\big). \tag{41}$$

Furthermore, setting $(s, t) = \big(2, \frac{7}{2}\big)$ and (2,5) in Lemma 3 yields $\|\boldsymbol{\mu}\|_{11/4}^{11/2} \le \|\boldsymbol{\mu}\|_2^2\|\boldsymbol{\mu}\|_{7/2}^{7/2}$ and $\|\boldsymbol{\mu}\|_{7/2}^7 \le \|\boldsymbol{\mu}\|_2^2\|\boldsymbol{\mu}\|_5^5$, from which it follows that

$$\widetilde{B}_{12}^{(1)} \le \widetilde{B}_{12}^{(4)}, \qquad \widetilde{B}_{12}^{(3)} \le \widetilde{B}_{12}^{(6)}. \tag{42}$$

Combining (40)–(42) with (39), we thus have

$$\widetilde{B}_{12} \le C\|\boldsymbol{\mu}\|_2\big(\widetilde{B}_{12}^{(4)} + \widetilde{B}_{12}^{(6)}\big) = C\|\boldsymbol{\mu}\|_2^3\big(A_4^2 + A_5^2\big). \tag{43}$$

Hence, by (36)–(38) and (43) we arrive at $\widetilde{B}_1 + \widetilde{B}_2 \le C\|\boldsymbol{\mu}\|_2^3\big(A_1^2 + A_2^2 + A_4^2 + A_5^2\big)$. This, together with (31), proves (25), and thus completes the proof of Theorem 1. □

To prove Theorem 2, we introduce the following auxiliary lemma, which provides a reverse Cauchy–Schwarz inequality for positive vectors.

Lemma 4. *Let $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$ be positive vectors (i.e. all entries are positive). For any real $s, t > 0$,*

$$\|\boldsymbol{x}\|_s^s\,\|\boldsymbol{y}\|_t^t \le \frac{x_{\max}^{s/2}\,y_{\max}^{t/2}}{x_{\min}^{s/2}\,y_{\min}^{t/2}}\left(\sum_{i=1}^n x_i^{s/2}y_i^{t/2}\right)^2.$$

Proof. Since all entries are positive, we have

$$x_{\min}^{s/2}y_{\min}^{t/2}\|\boldsymbol{x}\|_s^s\|\boldsymbol{y}\|_t^t = x_{\min}^{s/2}y_{\min}^{t/2}\sum_{i=1}^n\sum_{j=1}^n x_i^s y_j^t = \sum_{1\le i,j\le n} x_{\min}^{s/2}x_i^{s/2}x_i^{s/2}\cdot y_{\min}^{t/2}y_j^{t/2}y_j^{t/2}$$
$$\le \sum_{1\le i,j\le n} x_j^{s/2}x_i^{s/2}x_{\max}^{s/2}\cdot y_i^{t/2}y_j^{t/2}y_{\max}^{t/2}$$
$$= x_{\max}^{s/2}y_{\max}^{t/2}\left(\sum_{i=1}^n x_i^{s/2}y_i^{t/2}\right)^2,$$

which proves the inequality as claimed. □

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press

Applying Lemma 4 to $\boldsymbol{x} = \boldsymbol{y} = \boldsymbol{\mu}$ yields that, for any $s, t > 0$,

$$\|\boldsymbol{\mu}\|_s^s \, \|\boldsymbol{\mu}\|_t^t \le \left(\frac{\mu_{\max}}{\mu_{\min}}\right)^{(s+t)/2} \|\boldsymbol{\mu}\|_{(s+t)/2}^{s+t}. \tag{44}$$

Proof of Theorem 2. Since convergence in Kolmogorov distance implies convergence in distribution, Theorem 1 reduces our task to proving that under the conditions of Theorem 2,

$$A_\ell = o\Big(\|\boldsymbol{\mu}\|_2^{5/2}\|\boldsymbol{\mu}\|_3^6 + \|\boldsymbol{\mu}\|_2^{9/2}\Big), \quad \ell = 1, 2, \ldots, 5. \tag{45}$$

This relation holds for A_3 and A_4 only under the sparsity condition (4). Indeed, by (5) we have

$$A_3 = \|\boldsymbol{\mu}\|_2^{5/2}\|\boldsymbol{\mu}\|_5^5 = o\Big(\|\boldsymbol{\mu}\|_2^{9/2}\Big), \qquad A_4 = \|\boldsymbol{\mu}\|_2^{3/2}\|\boldsymbol{\mu}\|_{5/2}^{5/2}\|\boldsymbol{\mu}\|_5^{5/2} = o\Big(\|\boldsymbol{\mu}\|_2^{9/2}\Big),$$

and thus (45) holds for $\ell = 3$ and 4.

In the following, we verify (45) for A_1, A_2, and A_5 under either (8) or (9). We first assume (4) and (8).

Consider A_1. Noting that $\|\boldsymbol{\mu}\|_{3/2}^{3/4} = O\big(\|\boldsymbol{\mu}\|_2^3\big)$ under (8) and $\|\boldsymbol{\mu}\|_{5/2}^{5/4} = o\big(\|\boldsymbol{\mu}\|_2\big)$ under (4), we obtain

$$A_1 = \|\boldsymbol{\mu}\|_{3/2}^{3/4}\|\boldsymbol{\mu}\|_{5/2}^{5/4} = o\big(\|\boldsymbol{\mu}\|_2^4\big) = o\Big(\|\boldsymbol{\mu}\|_2^{9/2}\Big),$$

and thus (45) holds for $\ell = 1$.

Similarly, using the facts $\|\boldsymbol{\mu}\|_{7/2}^{7/4} = o(\|\boldsymbol{\mu}\|_3^{3/2})$ and $\|\boldsymbol{\mu}\|_4^2 = o(\|\boldsymbol{\mu}\|_3^{3/2})$, for A_2 we have

$$A_2 = \|\boldsymbol{\mu}\|_{3/2}^{3/4}\|\boldsymbol{\mu}\|_2^{1/2}\|\boldsymbol{\mu}\|_{7/2}^{7/4}\|\boldsymbol{\mu}\|_4^2 = o\Big(\|\boldsymbol{\mu}\|_2^{7/2}\|\boldsymbol{\mu}\|_3^3\Big) = o\Big(\Big(\|\boldsymbol{\mu}\|_2^{5/2}\|\boldsymbol{\mu}\|_3^6\Big)^{1/2}\Big(\|\boldsymbol{\mu}\|_2^{9/2}\Big)^{1/2}\Big).$$

Then it follows from the basic inequality $\sqrt{xy} \le (x+y)/2$ for $x, y > 0$ that (45) holds for $\ell = 2$.

For A_5, Lemma 3 with $s = \frac{3}{2}$ and $t = 2$ yields $\|\boldsymbol{\mu}\|_{7/4}^{7/2} \le \|\boldsymbol{\mu}\|_{3/2}^{3/2}\|\boldsymbol{\mu}\|_2^2$, which implies that

$$A_5 = \|\boldsymbol{\mu}\|_{7/4}^{7/4}\|\boldsymbol{\mu}\|_{7/2}^{7/4} \le \|\boldsymbol{\mu}\|_{3/2}^{3/4}\|\boldsymbol{\mu}\|_2\|\boldsymbol{\mu}\|_{7/2}^{7/4}. \tag{46}$$

Thus, by (8) and the fact that $\|\boldsymbol{\mu}\|_{7/2}^{7/4} = o(\|\boldsymbol{\mu}\|_3^{3/2})$, we have

$$A_5 = o\Big(\|\boldsymbol{\mu}\|_2^4\|\boldsymbol{\mu}\|_3^{3/2}\Big) = o\Big(\Big(\|\boldsymbol{\mu}\|_2^{5/2}\|\boldsymbol{\mu}\|_3^6\Big)^{1/4}\Big(\|\boldsymbol{\mu}\|_2^{9/2}\Big)^{3/4}\Big). \tag{47}$$

Then it follows from the inequality $(xy^3)^{1/4} \le (x+3y)/4$ for $x, y > 0$ that (45) holds for $\ell = 5$.

Finally, under (4) and (9), we invoke (44) to obtain

$$\|\boldsymbol{\mu}\|_s^{s/2}\|\boldsymbol{\mu}\|_t^{t/2} = O\Big(\|\boldsymbol{\mu}\|_2^{3(s+t)/8}\|\boldsymbol{\mu}\|_{(s+t)/2}^{(s+t)/2}\Big). \tag{48}$$

We also consider A_1, A_2, and A_5 in turn.

For $s = \frac{3}{2}$ and $t = \frac{5}{2}$, (48) gives

$$A_1 = \|\boldsymbol{\mu}\|_{3/2}^{3/4}\|\boldsymbol{\mu}\|_{5/2}^{5/4} = O\Big(\|\boldsymbol{\mu}\|_2^{7/2}\Big) = o\Big(\|\boldsymbol{\mu}\|_2^{9/2}\Big), \tag{49}$$

and thus (45) holds for $\ell = 1$.

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press

For A_2, by the second equality in (49) and the simple fact that $\|\boldsymbol{\mu}\|_4^2 = o\big(\|\boldsymbol{\mu}\|_3^{3/2}\big)$, we have

$$A_2 = \|\boldsymbol{\mu}\|_{3/2}^{3/4}\|\boldsymbol{\mu}\|_2^{1/2}\|\boldsymbol{\mu}\|_{7/2}^{7/4}\|\boldsymbol{\mu}\|_4^2 = o\Big(\|\boldsymbol{\mu}\|_{3/2}^{3/4}\|\boldsymbol{\mu}\|_2^{1/2}\|\boldsymbol{\mu}\|_{5/2}^{5/4}\|\boldsymbol{\mu}\|_4^2\Big) = o\Big(\|\boldsymbol{\mu}\|_2^4\|\boldsymbol{\mu}\|_3^{3/2}\Big),$$

which matches (47), and thus (45) holds for $\ell = 2$.

Consider A_5. Note that (46) holds under (4), and Lemma 3 with $s = 2$ and $t = 3$ gives $\|\boldsymbol{\mu}\|_{5/2}^{5/2} \le \|\boldsymbol{\mu}\|_2\|\boldsymbol{\mu}\|_3^{3/2}$. Applying (48) with $s = \frac{3}{2}$ and $t = \frac{7}{2}$ to (46), we thus have

$$A_5 = o\Big(\|\boldsymbol{\mu}\|_2^{23/8}\|\boldsymbol{\mu}\|_{5/2}^{5/2}\Big) = o\Big(\|\boldsymbol{\mu}\|_2^3\|\boldsymbol{\mu}\|_{5/2}^{5/2}\Big) = o\Big(\|\boldsymbol{\mu}\|_2^4\|\boldsymbol{\mu}\|_3^{3/2}\Big),$$

which also matches (47). This proves (45) for $\ell = 5$, and completes the proof of Theorem 2. □

Acknowledgements

We wish to thank two anonymous referees for their constructive comments that helped improve the quality of the paper.

Funding information

There are no funding bodies to thank relating to the creation of this article.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] Abbe, E. (2018). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18**, 1–86.
[2] Alon, U. (2007). Network motifs: Theory and experimental approaches. *Nature Rev. Genet.* **8**, 450–461.
[3] Barabási, A. L. and Albert, R. (1999). Emergence of scaling in random networks. *Science* **286**, 509–512.
[4] Barbour, A. D. and Chen, L. H. Y. (2005). *An Introduction to Stein's Method*. World Scientific, Singapore.
[5] Bhattacharya, B. B., Chatterjee, A. and Janson, S. (2023). Fluctuations of subgraph counts in graphon based random graphs. *Combinatorics Prob. Comput.* **32**, 428–464.
[6] Bollobás, B. (2001). *Random Graphs*, 2nd edn. Cambridge University Press.
[7] Bullmore, E. and Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Rev. Neurosci.* **10**, 186–198.
[8] Butzek, M. and Eichelsbacher, P. (2024). Non-uniform Berry–Esseen bounds for Gaussian, Poisson and Rademacher processes. Preprint, arXiv:2409.09439v1.
[9] Chang, J., Hu, Q., Kolaczyk, E. D., Yao, Q. and Yi, F. (2024). Edge differentially private estimation in the β-model via jittering and method of moments. *Ann. Statist.* **52**, 508–528.
[10] Chatterjee, S., Diaconis, P. and Sly, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Prob.* **21**, 1400–1435.
[11] Chen, M., Kato, K. and Leng, C. (2021). Analysis of networks via the sparse β-model. *J. R. Statist. Soc. B* **83**, 887–910.
[12] Clauset, A., Shalizi, C. R. and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Rev.* **51**, 661–703.
[13] Eichelsbacher, P. and Rednoss, B. (2023). Kolmogorov bounds for decomposable random variables and subgraph counting by the Stein–Tikhomirov method. *Bernoulli* **29**, 1821–1848.
[14] Eichelsbacher, P., Rednoss, B., Thäle, C. and Zheng, G. (2023). A simplified second-order Gaussian Poincaré inequality in discrete setting with applications. *Ann. Inst. H. Poincaré Prob. Statist.* **59**, 271–302.
[15] Estrada, E. (2010). Quantifying network heterogeneity. *Phys. Rev. E* **82**, 066102.
[16] Hladký, J., Pelekis, C. and Šileikis, M. (2021). A limit theorem for small cliques in inhomogeneous random graphs. *J. Graph Theory* **97**, 578–599.

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press

[17] HOLLAND, P. W. AND LEINHARDT, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Stat. Assoc.* **76**, 33–50.
[18] JIN, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43**, 57–89.
[19] KARRER, B. AND NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83**, 016107.
[20] KARWA, V. AND SLAVKOVIĆ, A. (2016). Inference using noisy degrees: Differentially private β-model and synthetic graphs. *Ann. Statist.* **44**, 87–112.
[21] KROKOWSKI, K., REICHENBACHS, A. AND THÄLE, C. (2016). Berry–Esseen bounds and multivariate limit theorems for functionals of Rademacher sequences. *Ann. Inst. H. Poincaré Prob. Statist.* **52**, 763–803.
[22] KROKOWSKI, K., REICHENBACHS, A. AND THÄLE, C. (2017). Discrete Malliavin–Stein method: Berry–Esseen bounds for random graphs and percolation. *Ann. Prob.* **45**, 1071–1109.
[23] MILO, R., SHEN-ORR, S., ITZKOVITZ, S., KASHTAN, N., CHKLOVSKII, D. AND ALON, U. (2002). Network motifs: Simple building blocks of complex networks. *Science* **298**, 824–827.
[24] MOSSEL, E., NEEMAN, J. AND SLY, A. (2015). Reconstruction and estimation in the planted partition model. *Prob. Theory Relat. Fields* **162**, 431–461.
[25] NEWMAN, M. E. J. (2009). Random graphs with clustering. *Phys. Rev. Lett.* **103**, 058701.
[26] NEWMAN, M. E. J. (2018). *Networks*, 2nd edn. Oxford University Press.
[27] NEWMAN, M. E. J., STROGATZ, S. H. AND WATTS, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 026118.
[28] NOURDIN, I. AND PECCATI, G. (2012). *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality*. Cambridge University Press.
[29] NOURDIN, I., PECCATI, G. AND REINERT, G. (2010). Stein's method and stochastic analysis of Rademacher functionals. *Electron. J. Prob.* **15**, 1703–1742.
[30] NUALART, D. (2006). *The Malliavin Calculus and Related Topics*. Springer, Berlin.
[31] PRIVAULT, N. (2008). Stochastic analysis of Bernoulli processes. *Prob. Surv.* **5**, 435–483.
[32] PRIVAULT, N. AND SERAFIN, G. (2020). Normal approximation for sums of weighted U-statistics – application to Kolmogorov bounds in random subgraph counting. *Bernoulli* **26**, 587–615.
[33] RIBEIRO, P., PAREDES, P., SILVA, M. E. P., APARICIO, D. AND SILVA, F. (2021). A survey on subgraph counting: Concepts, algorithms, and applications to network motifs and graphlets. *ACM Comput. Surv*. **54**, Article 28.
[34] RINALDO, A., PETROVIĆ, S. AND FIENBERG, S. E. (2013). Maximum likelihood estimation in the β-model. *Ann. Statist.* **41**, 1085–1110.
[35] ROBINS, G., PATTISON, P., KALISH, Y. AND LUSHER, D. (2007). An introduction to exponential random graph models for social networks. *Soc. Networks* **29**, 173–191.
[36] RÖLLIN, A. (2022). Kolmogorov bounds for the normal approximation of the number of triangles in the Erdös–Rényi random graph. *Prob. Eng. Inf. Sci.* **36**, 747–773.
[37] RUCIŃSKI, A. (1988). When are small subgraphs of a random graph normally distributed? *Prob. Theory Relat. Fields* **78**, 1–10.
[38] SVANTE, J., ŁUCZAK, T. AND RUCIŃSKI, A. (2000). *Random Graphs*. John Wiley, New York.
[39] TEIXEIRA, C. H. C., FONSECA, A. J., SERAFINI, M., SIGANOS, G., ZAKI, M. J. AND ABOULNAGA, A. (2015). Arabesque: A system for distributed graph mining. In *Proc. 25th Symp. Operating Systems Principles*. ACM, New York, pp. 425–440.
[40] VAN DER HOFSTAD, R., VAN DER HOORN, P., LITVAK, N. AND STEGEHUIS, C. (2020). Limit theorems for assortativity and clustering in null models for scale-free networks. *Adv. Appl. Prob.* **52**, 1035–1084.
[41] WATTS, D. J. AND STROGATZ, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442.
[42] YAN, T. AND XU, J. (2013). A central limit theorem in the inline316-model for undirected random graphs with a diverging number of vertices. *Biometrika* **100**, 519–524.

https://doi.org/10.1017/jpr.2026.10090 Published online by Cambridge University Press