Submitted to the Annals of Statistics

EDGEWORTH EXPANSIONS FOR NETWORK MOMENTS

BY YUAN ZHANG[1] AND DONG XIA[2]

[1]*Department of Statistics, The Ohio State University, yzhanghf@stat.osu.edu*

[2]*Department of Mathematics, The Hong Kong University of Science and Technology, madxia@ust.hk*

Network method of moments [20] is an important tool for nonparametric network inference. However, there has been little investigation on accurate descriptions of the sampling distributions of network moment statistics. In this paper, we present the first higher-order accurate approximation to the sampling CDF of a studentized network moment by Edgeworth expansion. In sharp contrast to classical literature on *noiseless* U-statistics, we show that the Edgeworth expansion of a network moment statistic as a *noisy* U-statistic can achieve higher-order accuracy without non-lattice or smoothness assumptions but just requiring weak regularity conditions. Behind this result is our surprising discovery that the two typically-hated factors in network analysis, namely, sparsity and edge-wise observational errors, jointly play a blessing role, contributing a crucial *self-smoothing* effect in the network moment statistic and making it analytically tractable. Our assumptions match the minimum requirements in related literature. For sparse networks, our theory shows that our empirical Edgeworth expansion and a simple normal approximation both achieve the same gradually depreciating Berry-Esseen type bound as the network becomes sparser. This result also significantly refines the best previous theoretical result.

For practitioners, our empirical Edgeworth expansion is highly accurate and computationally efficient. It is also easy to implement and convenient for parallel computing. We demonstrate the clear advantage of our method by several comprehensive simulation studies. As a byproduct, we also provide a finite-sample analysis of the network jackknife.

We showcase three applications of our results in network inference. We prove, to our knowledge, the first theoretical guarantee of higher-order accuracy for some network bootstrap schemes, and moreover, the first theoretical guidance for selecting the sub-sample size for network sub-sampling. We also derive a one-sample test and the Cornish-Fisher confidence interval for a given moment with higher-order accurate controls of confidence level and type I error, respectively.

1. Introduction.

1.1. *Overview.* *Network moments* are the frequencies of particular patterns, called *motifs*, that repeatedly occur in networks [102, 7, 114]. Examples include triangles, stars and wheels. They provide succinct and informative sketches of potentially very high-dimensional network population distributions. Pioneered by [20, 95], the *method of moments* for network data has become a powerful tool for frequentist nonparametric network inferences [8, 101, 131, 6, 99]. Compared to model-based network inference methods [91, 128, 94], moment method enjoys several unique advantages.

First, network moments play important roles in network modeling. They are the building blocks of the well-known exponential random graph models (ERGM) [78, 135]. More generally, under an exchangeable network assumption, the deep theory by [20] (Theorem 3) and [26] (Theorem 2.1) show that knowing all population moments can uniquely determine the network model up to weak isomorphism, despite no explicit inversion formula is yet available. From the perspective of statistical inference, evaluation of network moments

MSC2020 subject classifications: Primary 62E17, 91D30; secondary 60F05.

Keywords and phrases: network inferences, method of moments, Edgeworth expansion, noisy U-statistic, network bootstrap, network jackknife.

is completely model-free, making them objective evidences for specification, validation and comparison of network models [27, 117, 125, 106]. Second, network moments can be very efficiently computed, easily allowing parallel computing. This is a crucial advantage in a big data era, where business and industry networks could contain $10^5 \sim 10^7$ or even more nodes [43, 92] and computation efficiency becomes a substantive practicality concern. Model-fitting based network inference methods might face challenges in handling huge networks, while moment method equipped with proper sampling techniques [112, 46] will scale more comfortably (also see our comment in Section 6). Third, many network moments and their derived functionals are important structural features of great practical interest. Examples include clustering coefficient [76, 130], degree distribution [109, 122], transitivity [113], and more listed in Table A.1 in [114].

Despite the importance and raising interest in network moment method, the answer to the following core question remains under-explored:

What is the sampling distribution of a network moment?

For a given network motif R[1], let $\widehat{U}_n$ denote its sample relative frequency (see (2.3) for a formal definition) with expectation $\mu_n := \mathbb{E}[\widehat{U}_n]$. Let $\widehat{S}_n^2$ be an estimator of $\mathrm{Var}(\widehat{U}_n)$ that we shall specify later. We are mainly interested in finding the distribution of the studentized form $\widehat{T}_n := (\widehat{U}_n - \mu_n)/\widehat{S}_n$. It is well-known that under the widely-studied *exchangeable network* model framework (see formal definition in Section 2.1), we have $\widehat{T}_n \xrightarrow{d} N(0,1)$ uniformly for "not too sparse" networks [20, 17, 61], but usually, $N(0,1)$ only provides a rough characterization of the CDF $F_{\widehat{T}_n}$, and one naturally yearns for a finer approximation. To this end, several network bootstrap methods have been recently proposed [20, 17, 61, 93] in an attempt to address this question. They quickly inspired many follow-up works [124, 123, 60, 37] that clearly reflect data analysts' need of an accurate approximation method. However, compared to their empirical effectiveness, the theoretical foundation of network bootstraps remains weak. Almost all existing justifications of network bootstraps critically depend on the following type of results

$$|\widehat{U}_n^* - \widehat{U}_n| = o_p(n^{-1/2}), \quad \text{and} \quad |\widehat{U}_n - U_n| = o_p(n^{-1/2});$$

$$\text{or similarly, } \left|\widehat{T}_n^* - \widehat{T}_n\right| = o_p(1), \quad \text{and} \quad |\widehat{T}_n - T_n| = o_p(1);$$

where $\widehat{U}_n^*$ or $\widehat{T}_n^*$ are bootstrapped statistics and U_n or T_n are noiseless versions (see formal definitions in Section 2.2). Then the validity of network bootstraps is implied by the well-known asymptotic normality of U_n or T_n [17, 61]. However, this approach cannot show whether network bootstraps have any accuracy advantage over a simple normal approximation, especially considering the much higher computational costs of bootstraps.

In this paper, we propose the first provable *higher-order accurate* approximation to the sampling distribution of a given studentized network moment. Our paper uncovers, for the first time, that in fact the noisy $\widehat{U}_n$ and $\widehat{T}_n$ are usually more analytically tractable than the noiseless versions U_n and T_n. This enables our original analysis that sharply contrasts the common approach in existing network bootstrap literature that studies $\widehat{U}_n$ by approximately reducing it to U_n.

Now, we briefly summarize our main results by an informal theorem here. Before presenting the main results, we make a few preparatory definitions.

[1] Without confusion, in this paper, we use R to represent both the motif as a subgraph pattern and its corresponding adjacency matrix representation.

DEFINITION 1.1 (Acyclic and cyclic motifs, see also [20, 17, 61, 93]). *A motif R is called* acyclic, *if its edge set is a subset of an r-tree. The motif is called* cyclic, *if it is* connected *and contains at least one cycle. In other words, a* cyclic *motif is connected but not a tree.*

DEFINITION 1.2. *To simplify the narration of our method's error bounds under different motif shapes, especially in Table 2 and proof steps, define the following shorthand*

$$\mathcal{M}(\rho_n, n; R) := \begin{cases} (\rho_n \cdot n)^{-1} \cdot \log^{1/2} n + n^{-1} \cdot \log^{3/2} n, & \text{For acyclic } R \\ \rho_n^{-r/2} \cdot n^{-1} \cdot \log^{1/2} n + n^{-1} \cdot \log^{3/2} n, & \text{For cyclic } R \end{cases} \tag{1.1}$$

To simplify the narration of tail-probability control, we define the following symbol.

DEFINITION 1.3. *For a sequence of random variables $\{Z_n\}$ and a deterministic sequence $\{\alpha_n\}$, define $\widetilde{O}_p(\cdot)$ as follows*

$$\text{We write } Z_n := \widetilde{O}_p(\alpha_n), \text{ if } \mathbb{P}(|Z_n| \geqslant C\alpha_n) = O(n^{-1}) \text{ for some constant } C > 0. \tag{1.2}$$

Our "$\widetilde{O}_p$" is similar to "o_p" in [96] (see the remark beneath its Lemma 2) and Assumption (A1) in [90]. For technical reasons, in this paper, we do not need to define a $\widetilde{o}_p(\cdot)$ sign.

Now we are ready to present the informal statement of our main results.

THEOREM 1.1 (Informal statement of main results). *Assume the network is generated by an exchangeable network model. Define the population Edgeworth expansion for a given network moment R with r nodes and s edges as follows:*

$$G_n(x) := \Phi(x) + \frac{\varphi(x)}{\sqrt{n} \cdot \xi_1^3} \cdot \left\{ \frac{2x^2+1}{6} \cdot \mathbb{E}[g_1^3(X_1)] \right.$$

$$\left. + \frac{r-1}{2} \cdot \left(x^2+1\right) \mathbb{E}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)] \right\},$$

where Φ and φ are the CDF and PDF of $N(0,1)$, respectively, and the estimable coefficients components ξ_1, $\mathbb{E}[g_1^3(X_1)]$ and $\mathbb{E}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)]$ will be defined in Section 3 and they only depend on the graphon f and the motif R. Let ρ_n denote the network sparsity parameter. For dense networks, under the assumptions:

1. $\rho_n^{-2s} \cdot \mathrm{Var}(g_1(X_1)) \geqslant$ *constant* > 0*;*
2. **(Dense regime)** $\rho_n = \omega(n^{-1/2})$ *for acyclic R, or* $\rho_n = \omega(n^{-1/r})$ *for cyclic R;*
3. *Either $\rho_n \preceq (\log n)^{-1}$, or* $\limsup_{t\to\infty} \left|\mathbb{E}\left[e^{\mathrm{i}tg_1(X_1)/\xi_1}\right]\right| < 1$*;*

we have

$$\left\| F_{\hat{T}_n}(u) - G_n(u) \right\|_\infty = O\left(\mathcal{M}(\rho_n, n; R)\right), \tag{1.3}$$

where $\|H(u)\|_\infty := \sup_{u\in\mathbb{R}} |H(u)|$, and $\mathcal{M}(\rho_n, n; R)$ (defined in (1.1)) satisfies $\mathcal{M}(\rho_n, n; R) \ll n^{-1/2}$. Under the same conditions, the empirical Edgeworth expansion $\widehat{G}_n$ with estimated coefficients (see (3.14)) satisfies

$$\left\| F_{\hat{T}_n}(u) - \widehat{G}_n(u) \right\|_\infty = \widetilde{O}_p(\mathcal{M}(\rho_n, n; R)). \tag{1.4}$$

for a large enough absolute constant C.

For sparse networks, we replace condition 2 by:

2'. **(Sparse regime)** $n^{-1} \prec \rho_n \leq n^{-1/2}$ *for acyclic* R*, or* $n^{-2/r} \prec \rho_n \leq n^{-1/r}$ *for cyclic* R*,*

The population Edgeworth expansion and a simple $N(0,1)$ *approximation both achieve the following Berry-Esseen bound*[2]*:*

$$\left\|F_{\widehat{T}_n}(u) - G_n(u)\right\|_\infty \asymp \left\|F_{\widehat{T}_n}(u) - \Phi(u)\right\|_\infty = O\left(\mathcal{M}(\rho_n, n; R)\right) \bigwedge o(1). \tag{1.5}$$

The empirical Edgeworth expansion achieves

$$\left\|F_{\widehat{T}_n}(u) - \widehat{G}_n(u)\right\|_\infty = \widetilde{O}_p\left(\mathcal{M}(\rho_n, n; R)\right) \bigwedge o_p(1).$$

That is, in the sparse regime, the empirical Edgeworth expansion has the same proved error rate bound as $N(0,1)$*.*

1.2. *Our contributions.* Our contributions are three-fold. First, we establish the first provably higher-order accurate distribution approximations for network moments (1.3) and provide the first finite-sample error rate guarantee. The results originated from our discovery of the surprisingly blessing roles that network noise and sparsity jointly play in this setting. Our work reveals a new dimension to the understanding of these two components in network analysis. Second, we propose a provably highly accurate and computationally efficient empirical Edgeworth approximation (1.4) for practical use. Our method not only enjoys a significantly improved error control than network bootstrap methods in existing literature, but also computes much faster. Third, our results enable accurate and fast nonparametric network inference procedures.

To understand the strength of our main results (1.3) and (1.4), notice that for dense networks (see Assumption (ii) of Lemma 3.1), we achieve *higher-order accuracy* in distribution approximation *without non-lattice or smoothness assumption*. To our best knowledge, the non-lattice assumption is universally required to achieve higher-order accuracy in all literature for similar settings. However, this assumption is violated by some popular network models such as stochastic block model, arguably one of the most important and widely-used network models. Waiving the graphon smoothness assumption makes our approach a powerful tool for model-free exploratory network analysis and analyzing networks with high complexity and irregularities, see our discussion in Section 3.4.

Apart from the first higher-order approximation for dense networks, for sparse networks, we also establish a novel modified Berry-Esseen bound (1.5) for both our method and normal approximation – this is also the sharpest result to date. These results significantly improve over the previous best known $o(1)$ bound in literature [20, 17, 61, 93] and fills a large blank in the big picture. As the network sparsity ρ_n declines from $n^{-1/2}$ towards n^{-1} for acyclic R, or from $n^{-1/r}$ towards $n^{-2/r}$ for cyclic R, our result reveals a gradually depreciating uniform error bound. In the boundary case, where $\rho_n = \omega(n^{-1})$ (acyclic), or $\rho_n = \omega(n^{-2/r})$ (cyclic), our result matches the uniform consistency result in classical literature.

The key insight of our method is to view the sample network moment $\widehat{U}_n$ as a *noisy U-statistic*, where "noise" refers to edge-wise observational errors in the adjacency matrix A. Our analysis reveals the connection and differences between the noisy and the conventional *noiseless* U-statistic settings. We discover the surprisingly blessing roles that the two typically-hated factors, namely, *edge-wise observational errors* and *network sparsity* jointly play in this setting, roughly summarized by the following intuitions:

[2]Berry-Esseen bound for an asymptotically normal random variable $Y_n \xrightarrow{d} N(\mu, \sigma^2)$ refers to the finite error bound τ_n such that $\|F_{Y_n}(u) - F_{N(\mu,\sigma^2)}(u)\|_\infty \leqslant \tau_n$. This bound is typically discussed for CLT where Y_n is a centered and rescaled sample mean. Berry-Esseen bound for U-statistics: see [30].

1. The edge-wise errors behave like a smoother that tames potential distribution discontinuity due to a lattice or discrete network population[3];
2. Network sparsity elevates the smoothing effect of the observational error term to a sufficient level, such that $F_{\widehat{T}_n}$ becomes analytically tractable.

At first sight, the smoothing effect of edge-wise errors is rather counter-intuitive. For instance, generating a binary A from the probability matrix W is *discretizing* the edge probabilities drawn from a *continuum* $[0,1]$ into *binary* entries. How could this eventually yield a smoothing effect? In Section 3.1, we present two simple examples to illustrate the intuitive reason. In our proofs, we present original analysis to carefully quantify the impact of such smoothing effect. Our analysis techniques are very different from those in network bootstrap papers [17, 61, 93]. Also, it seems unlikely that our assumptions can be substantially relaxed since they match the well-known minimum conditions in related settings in [89].

Our empirical Edgeworth expansion (1.4) is model-free, assuming only weak regularity conditions; has the sharpest finite-sample error bound guarantees to date; computes very fast, much more scalable than network bootstraps; and easily permits parallel computing.

We showcase three applications of our main results. We present the first proof of the higher-order accuracy of some mainstream network bootstrap techniques under certain conditions, which their original proposing papers did not prove. Our results also enable rich future works on accurate and computationally very efficient network inferences. We present two immediate applications to testing and Cornish-Fisher type confidence interval for network moments with explicit accuracy guarantees.

1.3. *Paper organization.* The rest of this paper is organized as follows. In Section 2, we formally set up the problem and provide a detailed literature review. In Section 3, we present our core ideas, derive the Edgeworth expansions and establish their uniform approximation error bounds. We discuss different versions of the studentization form. We also present our modified Berry-Esseen theorem for the sparse regime. In Section 4, we present three applications of our results: bootstrap accuracy, one-sample test, and one-sample Cornish-Fisher confidence interval. In Section 5, we conduct three simulations to evaluate the performance of our method from various aspects. Section 6 discusses interesting implications of our results and future work.

1.4. *Big-O and small-o notation system.* In this paper, we will make frequent references to the big-O and small-o notation system. We use the same definitions of $O(\cdot)$, $o(\cdot)$, $\Omega(\cdot)$ and $\omega(\cdot)$ as that in standard mathematical analysis, and the same $O_p(\cdot)$ and $o_p(\cdot)$ as that in probability theory. For two deterministic series a_n and b_n, we write $a_n \preceq b_n$ to stand for $a_n = O(b_n)$, $n \to \infty$; and use $a_n \prec b_n$ or $a_n \ll b_n$ to stand for $a_n = o(b_n)$, $n \to \infty$; similarly define $\succeq$, $\succ$ and $\gg$.

2. Problem set up and literature review.

2.1. *Exchangeable networks and graphon model.* The base model of this paper is exchangeable network model [49, 19]. Exchangeability describes the unlabeled nature of many networks in social, knowledge and biological contexts, where node indices do not carry meaningful information. It is a very rich family that contains many popular models as special cases, including the stochastic block model and its variants including degree-corrected

[3] More precisely speaking, such irregularity is jointly induced by both the network population distribution and the shape of the motif, but the former is usually the determining factor.

stochastic block model and overlapping memberships [4] [75, 141, 139, 140, 3, 83, 137, 82, 57], the configuration model [42, 103], latent space models [74, 62] and general smooth graphon models [41, 56, 136][5]. In this paper, we base our study on the following exchangeable network model called *graphon model*. The framework is closely related to the Aldous-Hoover representation for infinite matrices [5, 77]. Under a graphon model, the n nodes correspond to latent space positions $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}[0,1]$. Network generation is governed by a measurable latent graphon function $f(\cdot,\cdot) : [0,1]^2 \to [0,1]$, $f(x,y) = f(y,x)$ that encodes all structures. The edge probability between nodes (i,j) is

$$W_{ij} = W_{ji} := \rho_n \cdot f(X_i, X_j); \quad 1 \leqslant i < j \leqslant n, \tag{2.1}$$

where the sparsity parameter $\rho_n \in (0,1)$ absorbs the constant factor, and we fix $\int_{[0,1]^2} f(u,v)\mathrm{d}u\mathrm{d}v = \text{constant}$. We only observe the adjacency matrix A with conditionally independent edges:

$$A_{ij} = A_{ji} | W \sim \text{Bernoulli}(W_{ij}), \forall 1 \leqslant i < j \leqslant n. \tag{2.2}$$

The model defined by (2.1) and (2.2) has a well-known issue that both f and $\{X_1, \dots, X_n\}$ are only identifiable up to equivalence classes [34]. This may pose significant challenges for model-based network inference, especially those based on parameter estimations. On the other hand, network moments are permutation-invariant and thus clearly immune to this identification issue. This makes network moments attractive study objectives.

2.2. *Network moment statistics.* To formalize network moments, it is more convenient to first define the sample version and then the population version. Each network moment is indexed by the corresponding motif R. For simplicity, we focus on connected motifs. Slightly abusing notation, here let R represent the adjacency matrix of a motif with r nodes and s edges. For any r-node sub-network $A_{i_1,\dots,i_r}$ [6] of A, define

$$h(A_{i_1,\dots,i_r}) := \mathbb{1}_{[A_{i_1,\dots,i_r} \sqsupseteq R]}{}^{7}, \quad \text{for all } 1 \leqslant i_1 < \cdots < i_r \leqslant n, \tag{2.3}$$

Here, "$A_{i_1,\dots,i_r} \sqsupseteq R$" means there exists a permutation map $\pi : \{1,\dots,r\} \to \{1,\dots,r\}$, such that $A_{i_1,\dots,i_r} \geqslant R_\pi$, where the "$\geqslant$" is entry-wise and R_π is defined as $(R_\pi)_{ij} := R_{\pi(i)\pi(j)}$. Our definition of $h(A_{i_1,\dots,i_r})$ here corresponds to the "$Q(R)$" defined in [20]. One can similarly define

$$\widetilde{h}(A_{i_1,\dots,i_r}) := \mathbb{1}_{[A_{i_1,\dots,i_r} \cong R]}, \quad \text{for all } 1 \leqslant i_1 < \cdots < i_r \leqslant n, \tag{2.4}$$

where "$A_{i_1,\dots,i_r} \cong R$" means there exists a permutation map $\pi : \{1,\dots,r\} \to \{1,\dots,r\}$, such that $A_{i_1,\dots,i_r} = R_\pi$. The definition of $\widetilde{h}$ corresponds to the "$P(R)$" studied in [20, 17], and [61]. As noted by [20], each h can be explicitly expressed as a linear combination of $\widetilde{h}$ terms, and vice versa. Therefore, they are usually treated with conceptual equivalence in literature, and most existing papers would choose one of them to study. For technical cleanness, in this

[4] Here we adopt the convention of [3, 19, 1] and view community memberships and degree corrections as random samples from their respective fixed hyper-distributions. There is a distinct understanding that memberships and degree corrections are completely free unknown model parameters [59], which our study does not cover.

[5] Smooth graphon: we can simply think that a graphon is called "smooth" if $f(\cdot,\cdot)$ is a smooth function. In the rigorous definition, f is smooth if $f(\psi(\cdot), \psi(\cdot))$ is smooth under some measure-preserving map $\psi : [0,1] \to [0,1]$, see [19, 56, 136].

[6] We write $A_{i_1,\cdots,i_r}$ to denote the sub-matrix of A with rows and columns indexed by $\{i_1, \cdots, i_r\}$.

[7] Since we consider an arbitrary but fixed R throughout this paper, without causing confusion, we drop the dependency on R in symbols such as h to simplify notation.

paper we focus on h. We believe our analysis also applies to $\widetilde{h}$, but the analysis is much more complicated and we leave it to future work. Define the *sample network moment* as

$$(2.5)\qquad \widehat{U}_n := \frac{1}{\binom{n}{r}} \sum_{1\leqslant i_1<\cdots<i_r\leqslant n} h(A_{i_1,\ldots,i_r}),$$

Then we define the *sample-population version* and *population version* of $\widehat{U}_n$ to be $U_n := \mathbb{E}[\widehat{U}_n|W]$ and $\mu_n := \mathbb{E}[U_n] = \mathbb{E}[\widehat{U}_n]$, respectively. We refer to $\widehat{U}_n$ as the *noisy* U-statistic, and call $U_n := \binom{n}{r}^{-1}\sum_{1\leqslant i_1<\cdots<i_r\leqslant n} h(W_{i_1,\ldots,i_r}) = \binom{n}{r}^{-1}\sum_{1\leqslant i_1<\cdots<i_r\leqslant n} h(X_{i_1},\ldots,X_{i_r})$[8] the conventional *noiseless* U-statistic, where we define $h(W_{i_1,\ldots,i_r}) = \mathbb{E}[h(A_{i_1,\ldots,i_r})|W]$, thus $\mu_n = \mathbb{E}[h(X_1,\cdots,X_r)]$. Similar to the insight that studentization is key to achieve higher-order accurate approximations in the i.i.d. setting (Section 3.5 of [129]), we study

$$\widehat{T}_n := \frac{\widehat{U}_n - \mu_n}{\widehat{S}_n},$$

where $\widehat{S}_n$ will be specified later in (3.3) and (3.4). We can similarly standardize or studentize the noiseless U-statistic U_n by $\check{T}_n := (U_n-\mu_n)/\sigma_n$ and $T_n := (U_n-\mu_n)/S_n$, respectively, where $\sigma_n^2 := \mathrm{Var}(U_n)$ and S_n^2 is a $\sqrt{n}$-consistent estimator[9] for σ_n^2, for instance, a jackknife variance estimator for the noiseless U-statistic U_n, c.f. [71, 96].

2.3. *Edgeworth expansions for i.i.d. data and noiseless U-statistics.* Edgeworth expansion [51, 127] refines the central limit theorem. It is the supporting pillar in the justification of bootstrap's higher-order accuracy, while itself is of great independent interest. In this subsection, we review the literature on Edgeworth expansions for i.i.d. data and conventional noiseless U-statistics, due to their close connection. Under mild conditions, the one-term Edgeworth expansion for the sample mean of n i.i.d. $X_1,\ldots,X_n$ reads $F_{n^{1/2}(\bar{X}-\mathbb{E}[X_1])/\sigma_{X_1}}(u) = \Phi(u) - n^{-1/2}\cdot\mathbb{E}[X_1^3](u^2-1)\varphi(u)/(6\sigma_{X_1}^3) + O(n^{-1})$, where Φ and φ are the CDF and PDF of $N(0,1)$, respectively. Edgeworth terms of even higher orders can be derived [68] but are not meaningful in practice unless we know a few true population moments. The minimax rate for estimating $\mathbb{E}[X_1^3]$ is $O_p(n^{-1/2})$, so $O(n^{-1})$ is the best practical remainder bound for an Edgeworth expansion. For further references, see [18, 115, 16, 66, 67, 10] and textbooks [68, 47, 129].

The literature on Edgeworth expansions for U-statistics concentrates on the noiseless version. In early 1980's, [30, 79, 32] established the asymptotic normality of the standarized and the studentized U-statistics, respectively, both with $O(n^{-1/2})$ Berry-Esseen type bounds. Then [31, 21, 90] approximated degree-two (i.e. $r=2$) standardized U-statistics with an $o(n^{-1})$ remainder with known population moments, and [14] established an $O(n^{-1})$ bound under relaxed conditions for more general symmetric statistics. Later, [71, 110] studied empirical Edgeworth expansions (EEE) with estimated coefficients and established $o(n^{-1/2})$ bounds. For finite populations, [11, 24, 25, 23] established the earliest results, and we will use some of their results in our analysis of network bootstraps. An incomplete list of other notable works on Edgeworth expansions for noiseless U-statistics with various finite moment assumptions includes [13, 70, 80, 96, 15, 81].

[8]Here, without causing confusion, we slightly abused the notation of $h(\cdot)$, letting it take either W or X as its argument, noticing that W is determined by $X_1,\ldots,X_n$. To elucidate $h(W_{i_1,\ldots,i_r})$, we first explicitly re-express $h(A_{i_1,\ldots,i_r})$ as a polynomial of $A_{i_1,\ldots,i_r}$'s edges, then replace "A" by "W". For example, with $R=$ triangle, we have $h(W_{123}) = W_{12}W_{13}W_{23} = \rho_n^3 f(X_1,X_2)f(X_1,X_3)f(X_2,X_3)$. Notice that generally, $h(W_{i_1,\ldots,i_r}) \neq \mathbb{1}_{[W_{i_1,\ldots,i_r}\sqsupseteq R]}$.

[9]$\sqrt{n}$-consistency of S_n^2 means that $\sqrt{n}(S_n^2-\sigma_n^2) = o_p(1)$, see [17, 93] for definition.

2.4. *The non-lattice condition and lattice Edgeworth expansions in the i.i.d. setting.* A major assumption called the *non-lattice condition* is critical for achieving $o(n^{-1/2})$ accuracy in Edgeworth expansions and is needed by all results in the i.i.d. setting without oracle moment knowledge and all results for noiseless U-statistics, but this condition is clearly not required for an $O(n^{-1/2})$ accuracy bound[10]. A random variable X_1 is called *lattice*, if it is supported on $\{a + bk : k \in \mathbb{Z}\}$ for some $a, b \in \mathbb{R}$ where $b \neq 0$. General discrete distributions are "nearly lattice" [11]. A distribution is essentially *non-lattice* if it contains a continuous component. In many works, the non-lattice condition is replaced by the stronger Cramer's condition [45]:

$$\limsup_{t\to\infty} \left|\mathbb{E}\left[e^{\mathrm{i}tX_1}\right]\right| < 1.$$

For U-statistics, this condition is imposed on $g_1(X_1) := \mathbb{E}[h(X_1, \ldots, X_r)|X_1] - \mu_n$. Cramer's condition can be relaxed [9, 100, 119, 120] towards a non-lattice condition, but all existing relaxations come at the price of essentially depreciated error bounds [12]. Therefore, for simplicity, in Theorems 3.1 and 4.1, we use Cramer's condition to represent the non-lattice setting.

However, in network analysis, Cramer's condition may be a strong assumption, for the following reasons. First, it is violated by stochastic block model, a very popular and important network model. In a block model, $g_1(X_1)$ only depends on node 1's community membership, thus is discrete. Second, this condition is difficult to check in practice. Third, some smooth models may even induce a lattice $g_1(X_1)$ under certain motifs and a non-lattice $g_1(X_1)$ under a different motif. For example, under the graphon model $f(x, y) := 0.3 + 0.1 \cdot \mathbb{1}_{[x>1/2;y>1/2]} + 0.1 \sin(2\pi(x + y))$, $g_1(X_1)$ is lattice when R is an edge, but it is non-lattice when R is a triangle.

Next, we review existing treatments of Edgeworth expansion in the lattice case that will spark the key inspiration to our work. In current literature, in the lattice case, we could approximate the CDF of an i.i.d. sample mean at higher-order accuracy, where the lattice Edgeworth expansion would contain an order $n^{-1/2}$ jump function; whereas to our best knowledge, no analogous result exists for U-statistics. Available approaches can be categorized into two mainstreams: (1) adding an artificial error term to the sample mean to smooth out lattice-induced discontinuity [118, 89]; and (2) formulating the lattice version Edgeworth expansion with a jump function [118]. The seminal work [118] adds a uniform error of bandwidth $n^{-1/2}$, and by inverting its impact on the smoothed distribution function, it explicitly formulates the lattice Edgeworth expansion with an $O(n^{-1})$ remainder. Another classical work [89] uses a normal artificial error instead of uniform and shows that the Gaussian bandwidth must be $\omega((\log n/n)^{1/2})$ and $o(1)$ to provide sufficient smoothing effect without causing an $\omega(n^{-1/2})$ distribution distortion. Other notable works include [132, 86, 12], in which, [132] and [86] also formulate lattice Edgeworth expansions in the i.i.d. univariate setting, and [12] studies Edgeworth expansions for the sample mean of i.i.d. random vectors, where some dimensions are lattice and the others are non-lattice.

[10] Simply use a Berry-Esseen theorem.

[11] "A discrete distribution is nearly-lattice": a discrete distribution, if not already lattice, can be viewed as a lattice distribution with diminishing periodicity.

[12] To our knowledge, existing results assuming only non-latticeness achieve no better than $o(n^{-1/2})$ approximation errors. For example, [14] replaces the RHS "1" in Cramer's condition by $1 - q$ and assumes it holds for $t \le n^{1/2}$. They obtain an error bound proportional to q^{-2}. Another example is [25]. It replaces [14]'s t range by $t \le \pi$ (their π is a variable) and obtains an error bound proportional to $q^{-2}\pi^{-2}$. Also see the comment beneath equation (4.7) of [110].

Despite the significant achievements of these treatments, latticeness remains an obstacle in practice. The difficulties are two-fold. On one hand, if we introduce an artificial error to smooth the distribution, it will unavoidably bring an $\Omega(n^{-1/2})$ distortion to the original distribution[13]. On the other hand, the exact formulation of a lattice Edgeworth expansion contains an $n^{-1/2}$ jump term. In many examples such as bootstrap, the jump locations depend on the true population variance, laying an uncrossable $\Omega(n^{-1/2})$ barrier for practical CDF approximation. For more details, see page 91 of [68].

3. Edgeworth expansions for network moments. Our approach to formulate the Edgeworth expansion can be summarized into the following progressive steps. We naturally start with decomposing $\widehat{U}_n$ and study the stochastic variation of each term in its expansion. Based on this understanding, we can design $\widehat{S}_n^2$ to estimate $\mathrm{Var}(\widehat{U}_n)$, studentize $\widehat{U}_n$ and formulate $\widehat{T}_n := (\widehat{U}_n - \mu_n)/\widehat{S}_n$. But using $\widehat{S}_n$ on the denominator of $\widehat{T}_n$ introduces additional first order (i.e. $O(n^{-1/2})$) bias in the eventual distribution approximation formula and also alters the approximately-Gaussian error term that contributes the key *self-smoothing* effect. Bearing this in mind, we expand $\widehat{T}_n$ and study the impact of the terms in this decomposition. The outcome of this part of analysis is the Edgeworth expansion formula. We then present our main theoretical results on explicit uniform and finite-sample error bounds for population and sample Edgeworth expansions, for dense and sparse networks, respectively. We conclude this section by a comprehensive comparison table of our results to existing literature and further discussions on the assumptions and results of our theory.

3.1. *Decomposition of the stochastic variations of $\widehat{U}_n$ and design of the variance estimator $\widehat{S}_n^2$.* The starting point of all analysis is the decomposition of $\widehat{U}_n$. This would allow us to design a variance estimator of $\widehat{U}_n$ for studentization. The studentized form, $\widehat{T}_n$, has a related but different decomposition, which will be formulated and analyzed next in Section 3.2. Now let us inspect $\widehat{U}_n$.

The stochastic variations in $\widehat{U}_n - \mu_n = (U_n - \mu_n) + (\widehat{U}_n - U_n)$ stem from two sources: (1) the randomness in $U_n - \mu_n$ due to W and ultimately $X_1, \dots, X_n$; and (2) the randomness in $\widehat{U}_n - U_n$ due to $A|W$, the edge-wise observational errors. In $\mathrm{Var}(\widehat{U}_n) = \mathbb{E}[\mathrm{Var}(\widehat{U}_n|W)] + \mathrm{Var}(\mathbb{E}[\widehat{U}_n|W])$, by Lemma 3.1, we observe $\mathrm{Var}(\widehat{U}_n|W) \asymp \rho_n^{2s-1} \cdot n^{-2}$ and $\mathrm{Var}(\mathbb{E}[\widehat{U}_n|W]) = \mathrm{Var}(U_n) \asymp \rho_n^{2s} \cdot n^{-1}$. We shall universally assume $\rho_n \cdot n \to \infty$, so $\sigma_n^2 = \mathrm{Var}(U_n) = \mathrm{Var}(\mathbb{E}[\widehat{U}_n|W])$ dominates. Therefore, our design of the variance estimator $\widehat{S}_n^2$ for $\mathrm{Var}(\widehat{U}_n)$ should align with the formulation of $\mathrm{Var}(U_n - \mu_n)$.

Now we inspect the main term $U_n - \mu_n$. It is a conventional noiseless U-statistic that admits the well-known Hoeffding's decomposition [73]:

$$(3.1)\quad U_n - \mu_n = \underbrace{\frac{r}{n}\sum_{i=1}^{n} g_1(X_i)}_{\text{Linear part}} + \underbrace{\frac{r(r-1)}{n(n-1)}\sum_{1\leqslant i<j\leqslant n} g_2(X_i, X_j)}_{\text{Quadratic part}} + \underbrace{\tilde{O}_p(\rho_n^s \cdot n^{-3/2}\log^{3/2} n)}_{\text{Higher-degree part}}$$

where $g_1, \dots, g_r$ are defined as follows. To avoid complicated subscripts, without confusion we define g_k's for special indexes $(i_1, \dots, i_r) = (1, \dots, r)$. For indexes 1 and $k \in \{2, \dots, r-1\}$ (only when $r \geqslant 3$) and r, define $g_1(x_1) := \mathbb{E}[h(X_1, \dots, X_r)|X_1 = x_1] - \mu_n$, $g_k(x_1, \dots, x_k) := \mathbb{E}[h(X_1, \dots, X_r)|X_1 = x_1, \dots, X_k = x_k] - \mu_n - \sum_{k'=1}^{k-1}\sum_{1\leqslant i_1<\dots<i_{k'}\leqslant r} g_{k'}(x_{i_1}, \dots, x_{i_{k'}})$ for $2 \leqslant k \leqslant r-1$ and $g_r(x_1, \dots, x_r) := h(x_1, \dots, x_r) - \mu_n$. From classical literature,

[13]To see this, simply notice that the original distribution contains $n^{-1/2}$ jumps, but the smoothed distribution does not, so an $o(n^{-1/2})$ approximation error is impossible [21].

we know that $\mathbb{E}[g_k(X_{i_1},\ldots,X_{i_k})|\{X_i : i \in \mathcal{I}_k \subset \{i_1,\ldots,i_k\}\}] = 0$, where the strict subset $\mathcal{I}_k$ could be $\varnothing$, and $\mathrm{Cov}\,(g_k(X_{i_1},\ldots,X_{i_k}), g_\ell(X_{j_1},\ldots,X_{j_\ell})) = 0$ unless $k = \ell$ and $\{i_1,\ldots,i_k\} = \{j_1,\ldots,j_\ell\}$. Consequently, the linear part in the Hoeffding's decomposition makes dominating contribution to $\mathrm{Var}(U_n - \mu_n)$[14]. Define

$$\xi_1^2 := \mathrm{Var}(g_1(X_1)). \tag{3.2}$$

Now we are ready to design $\widehat{S}_n$ and thus can fully specify $\widehat{T}_n = (\widehat{U}_n - \mu_n)/\widehat{S}_n$. There are two main choices of $\widehat{S}_n$. The conventional choice for studentizing noiseless U-statistics [32, 71, 110] uses the jackknife estimator

$$n \cdot \widehat{S}^2_{n;\text{jackknife}} := (n-1)\sum_{i=1}^n \left(\widehat{U}_n^{(-i)} - \widehat{U}_n\right)^2, \tag{3.3}$$

where $\widehat{U}_n^{(-i)}$ is $\widehat{U}_n$ calculated on the induced sub-network of A with node i removed. Despite conceptual straightforwardness, the jackknife estimator unnecessarily complicates analysis. In this paper, we propose an estimator with a simpler formulation. In $\mathrm{Var}(\widehat{U}_n) = \sigma_n^2 + O(\rho_n^{2s-1}n^{-2}) = r^2\xi_1^2/n + O(\rho_n^{2s-1}n^{-2})$, replace ξ_1 by its moment estimator. Specifically, recall that $\xi_1^2 = \mathrm{Var}(g_1(X_1)) = \mathbb{E}[(\mathbb{E}[h(X_1,\ldots,X_n)|X_1] - \mu_n)^2]$. Replacing $\mathbb{E}[h(X_1,\ldots,X_n)|X_1]$ and μ_n by their estimators based on observable data, we can design $\widehat{S}_n$ as follows

$$n \cdot \widehat{S}_n^2 := \frac{r^2}{n}\underbrace{\sum_{i=1}^n \left\{ \frac{1}{\binom{n-1}{r-1}} \sum_{\substack{1\leqslant i_1<\cdots<i_{r-1}\leqslant n\\ i_1,\ldots,i_{r-1}\neq i}} h(A_{i,i_1,\ldots,i_{r-1}}) - \widehat{U}_n \right\}^2}_{\text{Estimates } \xi_1^2=\mathrm{Var}(g_1(X_1))}. \tag{3.4}$$

We will show in Theorem 3.3 that the $|\widehat{S}_n^2 - \widehat{S}^2_{n;\text{jackknife}}|$ is ignorable, but our estimator $\widehat{S}_n$ is computationally much more efficient than the jackknife estimator. See our discussion right following Theorem 3.3.

3.2. *Expansion of $\widehat{T}_n$ and self-smoothing phenomenon.* The studentization $\widehat{T}_n$ can be expanded using a similar method to our study of $\widehat{U}_n$, but certain into a very different expression. The analysis in Section 3.1 already gives us a good understanding of the expansion of $\widehat{T}_n$'s numerator, namely, recall that

$$\widehat{U}_n - \mu_n = \frac{r}{n}\sum_{i=1}^n g_1(X_i) + \frac{r(r-1)}{n(n-1)}\sum_{1\leqslant i<j\leqslant n} g_2(X_i,X_j) + (\widehat{U}_n - U_n) + \text{remainder} \tag{3.5}$$

where we shall prove that the remainder terms contributed by $g_k, k \geqslant 3$ are dominated by $\widehat{U}_n - U_n$. Now, to handle $\widehat{T}_n$'s denominator, we follow the method in Maesono [96] and re-express $\widehat{T}_n$ as:

$$\widehat{T}_n = \frac{\widehat{U}_n - \mu_n}{\widehat{S}_n} = \frac{\widehat{U}_n - \mu_n}{\sigma_n} \cdot \left\{1 + \frac{\widehat{S}_n^2 - \sigma_n^2}{\sigma_n^2}\right\}^{-1/2} \tag{3.6}$$

[14]Hoeffding's decomposition reveals that the asymptotic behavior of the noiseless U-statistic U_n is largely determined by the linear part and bears some similarity to the i.i.d. case. But we should also notice that the quadratic part, i.e. g_2 terms, plays a non-ignorable role in the Edgeworth expansion of U_n. For more details, see [21, 71, 96, 110]

and use Taylor expansion $(1+x)^{-1/2} \approx 1 - x/2 + O(x^2)$ with $x := (\widehat{S}_n^2 - \sigma_n^2)/\sigma_n^2 = \widetilde{O}_p(n^{-1/2})$. In fact, just like our earlier decomposition of $\widehat{U}_n - \mu_n$ into two parts that represent the random variations originated from W (or $X_1, \ldots, X_n$) and $A|W$, respectively; here, it is also technically beneficial to do the same for $\widehat{S}_n^2 - \sigma_n^2$. Define an auxiliary intermediate term $\widehat{\sigma}_n^2$ to insert in between $\widehat{S}_n^2$ and σ_n^2:

$$n \cdot \widehat{\sigma}_n^2 := \frac{r^2}{n} \sum_{i=1}^n \left\{ \frac{1}{\binom{n-1}{r-1}} \sum_{\substack{1 \leqslant i_1 < \cdots < i_{r-1} \leqslant n \\ i_1, \ldots, i_{r-1} \neq i}} h(W_{i, i_1, \ldots, i_{r-1}}) - U_n \right\}^2.$$

and also define the following convenience shorthand

$$U_n^{\#} := \frac{1}{\sqrt{n} \cdot \xi_1} \sum_{i=1}^n g_1(X_i), \quad \Delta_n := \frac{r-1}{\sqrt{n}(n-1)\xi_1} \sum_{1 \leqslant i < j \leqslant n} g_2(X_i, X_j), \tag{3.7}$$

$$\widehat{\Delta}_n := (\widehat{U}_n - U_n)/\sigma_n, \quad \delta_n := (\widehat{\sigma}_n^2 - \sigma_n^2)/\sigma_n^2, \quad \text{and} \quad \widehat{\delta}_n := (\widehat{S}_n^2 - \widehat{\sigma}_n^2)/\sigma_n^2,$$

Recall that in Section 3.1 we observed that $\sigma_n^2 := \mathrm{Var}(U_n) \asymp r^2 \xi_1^2 / n$. We now obtain the key expansion of $\widehat{T}_n$ as follows:

$$\begin{aligned} \widehat{T}_n &= \left(U_n^{\#} + \Delta_n + \widehat{\Delta}_n + \widetilde{O}_p(n^{-1} \log^{3/2} n) \right) \cdot \left(1 + \widehat{\delta}_n + \delta_n \right)^{-1/2} \\ &= \widetilde{T}_n + \check{\Delta}_n + \text{Remainder}, \end{aligned} \tag{3.8}$$

in which we define

$$\widetilde{T}_n := U_n^{\#} + \Delta_n - \frac{1}{2} U_n^{\#} \cdot \delta_n, \tag{3.9}$$

$$\check{\Delta}_n := \binom{n}{2}^{-1} \sum_{1 \leqslant i < j \leqslant n} \widehat{\Theta}_{ij} \cdot \eta_{ij}, \tag{3.10}$$

where we define $\eta_{ij} := A_{ij} - W_{ij}$, and the formal definition of $\widehat{\Theta}_{ij}$ is lengthy and sunk to Supplemental Material (see (8.19)). The gist is that $\widehat{\Theta}_{ij}$ is a function of W (thus all its randomness comes from $X_1, \ldots, X_n$) and does *not* depend on the conditional randomness in $A|W$, and also that $\widehat{\Theta}_{ij} \asymp \rho_n^{-1} \cdot n^{1/2}$. The term $\check{\Delta}_n$ encodes the "linear part" (linear in η_{ij}'s) of $\widehat{\Delta}_n$ (see Lemma 3.1-(c)). The remainder in (3.8) consists of the remainder terms from the two expansions of $U_n - \mu_n$ and $\widehat{U}_n - U_n$, respectively. We will show that the remainder is $\widetilde{O}_p(\mathcal{M}(\rho_n, n; R))$, where we recall the definition of $\widetilde{O}_p$ from Section 1.4.

To give readers a quick preview of the roles of the main constituent terms in the expansion of $\widehat{T}_n$, we present a summary table, see Table 1. The full quantitative justification of its contents will be provided soon in Lemma 3.1. Notice that despite smoother $\check{\Delta}_n$ is $\Omega(n^{-1/2})$, it does *not* distort any smooth order-$n^{-1/2}$ term in the Edgeworth expansion formula. Similar phenomenon is observed in the i.i.d. setting, see [118] (equation (2.8)) and [89] (Section 2.2).

Our decomposition (3.8) is a renaissance of the spirits of [118] and [89], but with the following crucial conceptual distinctions. First and most important, the error term $\check{\Delta}_n$ in our formula is *not* artificial, but a natural constituent component of $\widehat{T}_n$. Therefore, the smoother does *not* distort the objective distribution, that is, $\widehat{T}_n$ is *self-smoothed*. The second distinction lies in the bandwidth of the smoothing error term. Since the smoothing error terms in [118] and [89] are artificial, the user is at the freedom to choose these bandwidths. In our setting, the bandwidth of the smoothing term $(\rho_n \cdot n)^{-1/2}$ is not managed by the user, but governed by the network sparsity. Therefore, when Cramer's condition fails, we make the very mild

TABLE 1
Summary of the main components in $\widehat{T}_n$

Component	Order of std. dev.	Impacts Edgeworth formula	Smoothing effect
$U_n^{\#}$	1	Yes	No
$\Delta_n - \frac{1}{2}U_n^{\#}\cdot\delta_n$	$n^{-1/2}$	Yes	No
$\breve{\Delta}_n$	$(\rho_n\cdot n)^{-1/2}$	No	Yes
Remainder	$\widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)$	No	No

sparsity assumption that $\rho_n = O((\log n)^{-1})$ to ensure enough smoothing effect. This echoes the lower bound on Gausssian bandwidth in [89]. This upper bound can be easily enforced by a data pre-processing step. See our discussion in Section 6. We also need ρ_n to be lower bounded to effectively bound the remainder term, see Lemma 3.1-(b). Third, our error term $\breve{\Delta}_n$ is *dependent* on $\widetilde{T}_n$ through W. Last, the proof technique of [118] is inapplicable to our setting due to the quadratic part ($g_2(X_i,X_j)$ terms) in $\widetilde{T}_n$; and [89] obtains an $o(n^{-1/2})$ error bound[15], while we aim at stronger results under a more complicated U-statistic setting with degree-two terms. In our proofs, we carefully manage these challenges with original analysis.

A key difference between our *noisy* U-statistic setting and the conventional *noiseless* setting is carried by the $\breve{\Delta}_n$ term, which is unique to network data. Prior to our paper, the typical treatment in network bootstrap literature is to simply bound and ignore this component, such as Lemma 7 in [61]. In sharp contrast, by carefully quantifying the impact of $\breve{\Delta}_n$, we shall reveal its key smoothing effect by a refined analysis. Therefore, before advancing to the statement of our main lemma, we present two concrete examples to give the general audience an intuitive impression of the asymptotic orders of each constituent term in (3.10). For simplicity of illustration, in these examples, we would *standardize* $\widehat{U}_n$ using its *true* variance σ_n^2, rather than the estimator $\widehat{S}_n^2$. The impact of this simplification is that the expansion of the standardization would not have the $-(1/2)U_n^{\#}\cdot\delta_n$ term, and an altered $\widehat{\Theta}_{ij}$ at the same asymptotic order as the original $\widehat{\Theta}_{ij}$, and a different remainder term; but all these differences are non-essential for demonstrating our core ideas. For the moment, let us bear in mind that $\sigma_n \asymp \widehat{S}_n \asymp \rho_n^s\cdot n^{-1/2}$ by Lemma 3.1. We first study the simplest motif $R = $ Edge.

EXAMPLE 3.1. *Let R be an edge with $r=2$ and $s=1$, and $\widehat{U}_n$ is simply the sample edge density $\widehat{\rho}_n := \bar{A}$. By definition, all $h(A_{i_1,i_2}) - h(W_{i_1,i_2})$ terms are mutually conditionally independent given W. Then the asymptotic behavior of the self-smoother term is*

$$\frac{\widehat{U}_n - U_n}{\sigma_n}\Big|W \xrightarrow{d} N\left(0, \sigma^2_{\frac{\widehat{U}_n-U_n}{\sigma_n}|W} \asymp (\rho_n\cdot n)^{-1}\right)$$

at a uniform $O(\rho_n^{-1/2}\cdot n^{-1})$ Berry-Esseen CDF approximation error rate.

The next example shows that the key insight of Example 3.1 also applies to general motifs.

EXAMPLE 3.2. *Let R be a triangular motif with $r=3, s=3$, and $\widehat{U}_n$ is the empirical triangle frequency. We can decompose $\widehat{U}_n - U_n$ as follows:*

$$\frac{\widehat{U}_n - U_n}{\sigma_n} = \frac{1}{\binom{n}{3}}\sum_{1\leqslant i_1<i_2<i_3\leqslant n}\frac{\{h(A_{i_1,i_2,i_3}) - h(W_{i_1,i_2,i_3})\}}{\sigma_n}$$

[15]The $o(n^{-1/2})$ error bound in [89] holds on some $\mathfrak{B}\subset\mathbb{R}$ with "diminishing boundary", while our error bounds hold on the entire $\mathbb{R}$.

$$=\frac{1}{\binom{n}{3}}\sum_{1\leqslant i_1<i_2<i_3\leqslant n}\frac{(W_{i_1i_2}+\eta_{i_1i_2})(W_{i_1i_3}+\eta_{i_1i_3})(W_{i_2i_3}+\eta_{i_2i_3})-W_{i_1i_2}W_{i_1i_3}W_{i_2i_3}}{\sigma_n}$$

$$=\frac{1}{\binom{n}{3}}\left\{\sum_{\substack{1\leqslant i_1<i_2\leqslant n\\1\leqslant i_3\leqslant n\\i_3\neq i_1,i_2}}\frac{W_{i_1i_3}W_{i_2i_3}\eta_{i_1i_2}+W_{i_1i_2}\eta_{i_1i_3}\eta_{i_2i_3}}{\sigma_n}+\sum_{1\leqslant i_1<i_2<i_3\leqslant n}\frac{\eta_{i_1i_2}\eta_{i_1i_3}\eta_{i_2i_3}}{\sigma_n}\right\}$$

$$=\underbrace{\frac{1}{\binom{n}{2}}\sum_{1\leqslant i<j\leqslant n}\left(\frac{3\sum_{\substack{1\leqslant k\leqslant n\\k\neq i,j}}W_{ik}W_{jk}}{(n-2)\sigma_n}\right)\eta_{ij}}_{\textit{Linear part}}+\underbrace{\frac{1}{\binom{n}{3}}\sum_{\substack{1\leqslant i<j\leqslant n\\1\leqslant k\leqslant n\\k\neq i,j}}\frac{W_{ij}}{\sigma_n}\eta_{ik}\eta_{jk}}_{\textit{Quadratic part}}$$

$$+\underbrace{\frac{1}{\binom{n}{3}}\sum_{1\leqslant i<j<k\leqslant n}\frac{1}{\sigma_n}\eta_{ij}\eta_{ik}\eta_{jk}}_{\textit{Cubic part}}$$

where recall that we define $\eta_{ij}:=A_{ij}-W_{ij}$. Recall that we are conditioning on W, so $\sigma_n\asymp\rho_n^s\cdot n^{-1/2}$ is treated as a constant. The linear part is $\asymp\rho_n^{-1/2}\cdot n^{-1/2}$, the quadratic part is $\widetilde{O}_p(\rho_n^{-1}\cdot n^{-1}\log^{1/2}n)$ and the cubic part is $\widetilde{O}_p(\rho_n^{-3/2}\cdot n^{-1}\log^{1/2}n)$. We make two observations. First, the linear part in this example has the same asymptotic order as the linear part in Example 3.1. This is not a coincidence and will be formalized by Lemma 3.1-(b). In other words, regardless of the shape of R, the linear part in such decomposition always provides smoothing effect at the same magnitude. Second, different from Example 3.1, we now have higher-degree remainder consisting of products of quadratic and cubic η terms. The linear part nicely always dominates the quadratic part; but it only dominates the cubic part when $\rho_n=\omega(n^{-1/2}\log^{1/2}n)$.

For readers' convenience, we now link the terms in the two examples to items in Table 1. The entire $(\widehat{U}_n-U_n)/\sigma_n$ in Example 3.1 and the linear part of the expansion in Example 3.2 both map to $\check{\Delta}_n$ in Table 1; and the quadratic and cubic parts of the expansion in Example 3.2 correspond to the remainder part in Table 1.

Readers who are familiar with the martingale CLT (c.f. [69]) see immediately that the cubic part in Example 3.2 is also asymptotically normal and naturally question why our study would stick to ρ_n regimes such that this term is ignorable. In other words, when the network is very sparse that the cubic part dominates the linear part, can the asymptotic normality of the former take over the role of self-smoother? The reason why the cubic part is much more challenging to characterize than the linear part lies in its very slow convergence to its limiting normal distribution. In Example 3.2, the CDF of the linear part converges to its limiting distribution at a uniform rate of $O(\rho_n^{-1/2}\cdot n^{-1})$ (See (3.12) in our Lemma 3.1-(b)). In sharp contrast, the convergence rate of the cubic part as a martingale is much slower. The reported uniform convergence rate for martingale CLT across various different settings in literature are all significantly slower than $n^{-1/2}$, see [72, 64, 107, 29] and so on. This is not surprising considering the lack of independence between summands in the scenarios that martingale CLT addresses. Our discussion here does not disprove the possibility that a sharper analysis might show that the higher-degree η-product terms in fact can serve as the self-smoother, but required analysis might be difficult. Considering the already existing complexity of this paper, we simply control the stochastic magnitude of the cubic part in Example 3.2.

On the other hand, however, the asymptotic normality of the cubic part provides a unrigorous but helpful intuitive understanding of the $\log^{1/2}n$ factor in the first term of our error bound (1.2). If we roughly treat $Z_{\text{cubic}} :=$ the cubic part in Example 3.2 as normal, then $\mathbb{P}(|Z_{\text{cubic}}| > C(\operatorname{Var}(Z_{\text{cubic}})^{1/2} \cdot \log^{1/2} n)) = O(n^{-1})$ for a large enough constant C. The $\log^{3/2} n$ factor in the second term of (1.2) comes from a different source, namely, the tail probability control of $g_k : k \geqslant 3$ terms in the Hoeffding's decomposition of $U_n - \mu_n$ (not presented by Example 3.2) in a similar spirit.

The insights of the two examples will be generalized in part (b) of our main lemma below. When the network is sufficiently dense, among the expansion terms of $\widehat{U}_n - U_n$, the linear part dominates. Consequently, the overall contribution of the stochastic variations in $A|W$ approximates Gaussian at an $O(\rho_n^{-1/2} \cdot n^{-1})$ Berry-Esseen error rate. Now recalling the definition of acyclic and cyclic R shapes from Definition 1.1, the definition of $\mathcal{M}(\rho_n, n; R)$ from definition 1.2 in Section 1, and the definition of $\widetilde{O}_p$, we are ready to state our main lemma.

LEMMA 3.1. *Assume the following conditions hold:*

(i). $\rho_n^{-s} \cdot \xi_1 > C > 0$, *where* $C > 0$ *is a universal constant,*
(ii). $\rho_n = \omega(n^{-1})$ *for acyclic* R, *or* $\rho_n = \omega(n^{-2/r})$ *for cyclic* R,

We have the following results:

(a) $\dfrac{U_n - \mu_n}{\sigma_n} = U_n^{\#} + \Delta_n + \widetilde{O}_p(n^{-1} \cdot \log^{3/2} n)$,
(b) *We have*

$$\widehat{\Delta}_n = \frac{(\widehat{U}_n - U_n)}{\sigma_n} = \breve{\Delta}_n + \breve{R}_n,$$

where $\breve{\Delta}_n$ *and the remainder* $\breve{R}_n$ *satisfy*

$$\breve{R}_n = \widetilde{O}_p\left(\mathcal{M}(\rho_n, n; R)\right) \tag{3.11}$$

$$\left\| F_{\breve{\Delta}_n|W}(u) - F_{N(0,(\rho_n \cdot n)^{-1}\sigma_w^2)}(u) \right\|_\infty = \widetilde{O}_p\left(\rho_n^{-1/2} \cdot n^{-1}\right) \tag{3.12}$$

where the order control in (3.12) *is* $\widetilde{O}_p(\cdot)$ *rather than* $O(\cdot)$ *due to the randomness in* W. *The definition of* σ_w *is lengthy and formally stated in Section 7 in Supplemental Material. As* $n \to \infty$, *we have* $\sigma_w \overset{p}{\asymp} 1$.
(c) $\widehat{\delta}_n = \widetilde{O}_p\left(\mathcal{M}(\rho_n, n; R)\right)$,
(d) *We have*

$$\delta_n = \frac{1}{n} \sum_{i=1}^{n} \frac{g_1^2(X_i) - \xi_1^2}{\xi_1^2} + \frac{2(r-1)}{n(n-1)} \sum_{\substack{1 \leqslant \{i,j\} \leqslant n \\ i \neq j}} \frac{g_1(X_i) g_2(X_i, X_j)}{\xi_1^2} + \widetilde{O}_p(n^{-1} \cdot \log n).$$

Overall, Lemma 3.1 clarifies the asymptotic orders of the leading terms in the expansion of $\widehat{T}_n$. In fact, Lemma 3.1 has a parallel version for the jackknife $\widehat{S}_{n;\text{jackknife}}$ in view of Theorem 3.3, but we do not present it due to page limit. We spend the rest of this section on discussing the conditions and results of Lemma 3.1.

REMARK 3.1. *Assumption* (i) *is a standard non-degeneration assumption in literature. It is different from a smoothness assumption on graphon* f[16]. *A globally smooth Erdos-Renyi graphon leads to a degenerate* $g_1(X_1)$ *that* $\xi_1^2 = \operatorname{Var}(g_1(X_1)) = 0$. *In the degenerate setting,*

[16]Smooth graphon: f is called *smooth*, if there exists a measure-preserving map $\varrho : [0,1] \to [0,1]$ such that $f(\varrho(\cdot), \varrho(\cdot))$ is a smooth function. See [56, 136] for more details.

both the standardization/studentization and the analysis would be very different. Asymptotic results for $r = 2, 3$ motifs under an Erdos-Renyi graphon have been established by [54, 55]. Degenerate U-statistics are outside the scope of this paper.

REMARK 3.2. *We note that Lemma 3.1 only requires the weak assumption on ρ_n (see Assumption ii). This assumption matches the classical sparsity assumptions in network bootstrap literature [20, 17, 61]. Using Lemma 3.1, we prove a higher-order error bound of the Edgeworth expansion in Theorem 3.1 with a stronger density assumption; while in Theorem 3.4 on sparse networks, we prove a novel modified Berry-Esseen bound for the normal approximation. Both downstream theorems significantly improve over existing best results.*

REMARK 3.3. *Lemma 3.1-(a) and (d) are similar to results in classical literature on Edgeworth expansion for noiseless U-statistics [71, 96], but here we account for ρ_n. Parts (b) and (c) are new results unique to the network setting. Especially in the proof of part (b), we significantly refine the analysis of the randomness in $A|W$ in [17] and [61].*

3.3. *Population and empirical Edgeworth expansions for network moments.* In this subsection, we present our main theorems.

THEOREM 3.1 (Population network Edgeworth expansion). *Define*

$$G_n(x) := \Phi(x) + \frac{\varphi(x)}{\sqrt{n}\cdot\xi_1^3}\cdot\left\{\frac{2x^2+1}{6}\cdot\mathbb{E}[g_1^3(X_1)]\right.$$

$$\left.+\frac{r-1}{2}\cdot\left(x^2+1\right)\mathbb{E}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)]\right\}, \tag{3.13}$$

where $\Phi(x)$ and $\varphi(x)$ are the CDF and PDF of $N(0,1)$, respectively. Assume condition (i) of Lemma 3.1 hold, and replace condition (ii) by a stronger assumption that either R is acyclic and $\rho_n = \omega(n^{-1/2})$, or R is cyclic and $\rho_n = \omega(n^{-1/r})$. Additionally, assume either $\rho_n = O((\log n)^{-1})$ or Cramer's condition $\limsup_{t\to\infty}\left|\mathbb{E}\left[e^{\mathbb{i}tg_1(X_1)\cdot\xi_1^{-1}}\right]\right| < 1$ holds. We have

$$\left\|F_{\widehat{T}_n}(x) - G_n(x)\right\|_\infty = O\left(\mathcal{M}(\rho_n, n; R)\right).$$

REMARK 3.4. *The assumed ρ_n's upper bound in absence of Cramer's condition serves to sufficiently boost the smoothing power of $\breve{\Delta}_n$, quantified in Lemma 3.1-(3.12). This assumption seems minimal in presence of a lattice $g_1(X_1)$, since it corresponds to a normal smoother with variance $(\rho_n\cdot n)^{-1} = \Omega(\log n\cdot n^{-1})$. This matches the minimum standard deviation requirement $\Omega((\log n)^{1/2}\cdot n^{-1/2})$ in Remark 2.4 in [89] for the i.i.d. setting.*

In (3.13), the Edgeworth coefficients depend on true population moments. In practice, they need to be estimated from data. Define

$$\widehat{g}_1(X_i) := \frac{1}{\binom{n-1}{r-1}}\sum_{\substack{1\leqslant i_1<\ldots<i_{r-1}\leqslant n\\ i_1,\ldots,i_{r-1}\neq i}} h(A_{i,i_1,\ldots,i_{r-1}}) - \widehat{U}_n,$$

$$\widehat{g}_2(X_i,X_j) := \frac{1}{\binom{n-2}{r-2}}\sum_{\substack{1\leqslant i_1<\ldots<i_{r-2}\leqslant n\\ i_1,\ldots,i_{r-2}\neq i,j}} h(A_{i,j,i_1,\ldots,i_{r-2}}) - \widehat{U}_n - \widehat{g}_1(X_i) - \widehat{g}_1(X_j),$$

where we write "$\widehat{g}_1(X_i)$" rather than "$\widehat{g_1(X_i)}$" for cleanness. We stress that the evaluation of $\widehat{g}_1(X_i)$ and $\widehat{g}_2(X_i, X_j)$ does *not* require knowing the latent X_i, X_j. The Edgeworth coefficients can be estimated by

$$\widehat{\xi}_1^2 := \frac{n \cdot \widehat{S}_n^2}{r^2} = \frac{1}{n}\sum_{i=1}^n \widehat{g}_1^2(X_i), \quad \text{and} \quad \widehat{\mathbb{E}}\left[g_1^3(X_1)\right] := \frac{1}{n}\sum_{i=1}^n \widehat{g}_1^3(X_i),$$

$$\widehat{\mathbb{E}}\left[g_1(X_1)g_1(X_2)g_2(X_1,X_2)\right] := \frac{1}{\binom{n}{2}} \sum_{1\leqslant i<j\leqslant n} \widehat{g}_1(X_i)\widehat{g}_1(X_j)\widehat{g}_2(X_i,X_j).$$

THEOREM 3.2 (Empirical network Edgeworth expansion). *Define the empirical Edgeworth expansion as follows:*

$$\widehat{G}_n(x) := \Phi(x) + \frac{\varphi(x)}{\sqrt{n}\cdot\widehat{\xi}_1^3} \cdot \left\{ \frac{2x^2+1}{6} \cdot \widehat{\mathbb{E}}[g_1^3(X_1)] \right.$$

$$\left. + \frac{r-1}{2} \cdot \left(x^2+1\right) \widehat{\mathbb{E}}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)] \right\}, \tag{3.14}$$

Under the conditions of Theorem 3.1, we have

$$\left\| F_{\widehat{T}_n}(x) - \widehat{G}_n(x) \right\|_\infty = \widetilde{O}_p(\mathcal{M}(\rho_n, n; R)).$$

REMARK 3.5. *Another approach to estimate the unknown coefficients in Edgeworth expansion is bootstrap. The concentration of $\widehat{G}_n \to G_n$ should not be confused with the concentration $\widehat{G}_n^* \to \widehat{G}_n$, where $\widehat{G}_n^*$ is the expansion with bootstrap-estimated coefficients. See literature regarding the i.i.d. setting [71, 96]. In $\widehat{G}_n^* \to \widehat{G}_n$, the convergence rate is not a concern, because without constraining computation cost, one can let the number of bootstrap samples grow arbitrarily fast. Hence, establishing consistency would suffice for the analysis of $\widehat{G}_n^* \to \widehat{G}_n$, whereas our proof concerning $\widehat{G}_n \to G_n$ requires careful rate calculations.*

Next, we show that different choices of the variance estimators for studentization represent no essential discrepancy.

THEOREM 3.3 (Studentizing by a jackknife variance estimator (3.3)). *Define*

$$\widehat{T}_{n;\text{jackknife}} := \frac{\widehat{U}_n - \mu_n}{\widehat{S}_{n;\text{jackknife}}}.$$

Under the assumptions of Theorem 3.1, we have

$$|\widehat{S}_n - \widehat{S}_{n;\text{jackknife}}| = O(\widehat{S}_n \cdot n^{-1}), \tag{3.15}$$

$$\left\| F_{\widehat{T}_{n;\text{jackknife}}}(x) - G_n(x) \right\|_\infty = O\left(\mathcal{M}(\rho_n, n; R)\right),$$

$$\left\| F_{\widehat{T}_{n;\text{jackknife}}}(x) - \widehat{G}_n(x) \right\|_\infty = \widetilde{O}_p(\mathcal{M}(\rho_n, n; R)).$$

Theorem 3.3 states that on statistical properties, one does not need to differentiate between $\widehat{T}_n$ and $\widehat{T}_{n;\text{jackknife}}$. The evaluation of $\widehat{S}_{n;\text{jackknife}}$ costs $O(n^{r+1})$ time because each individual $\widehat{U}_n^{(-i)}$ costs $O(n^r)$; whereas our estimator $\widehat{S}_n$ costs $O(n^r)$. Our estimator also has a more convenient form for theoretical analysis.

3.4. *Remarks on non-smooth graphons.* Our results do not assume graphon smoothness or low-rankness. This aligns with the literature on noiseless U-statistics but sharply contrasts network inferences based on model parameter estimation such as [74, 91] and network bootstraps based on model estimation [61, 93]. Notice that the concept "non-smoothness" usually emphasizes "not assuming smoothness" rather than explicitly describing irregularity. It is a very useful tool for modeling networks with high structural complexity or unbalanced observations, examples include: (1) a small group of *outlier* nodes that behave differently from the main network patterns [28]; (2) in networks that exhibit "core-periphery" structures [48, 138], we may wish to relax structural assumptions on periphery nodes due to scarcity of observations; and (3) networks generated from a mixture model [104, 82] with many small-probability mixing components may appear non-smooth in these parts. Unfortunately, existing research on practical methods for non-smooth graphons is rather limited due to the obvious technical difficulty, but exceptions include [40].

Our results send the surprising message that under mild conditions, the sampling distribution of a network moment is still *smooth* and can be *accurately* approximated, even if the graphon is non-smooth.

3.5. *Sparse networks.* We have been focusing on discussing dense networks, but many networks tend to be sparse [63]. In this section, we investigate the following sparsity regime

$$\rho_n : \begin{cases} n^{-1} \prec \rho_n \preceq n^{-1/2}, & \text{for acyclic } R \\ n^{-2/r} \prec \rho_n \preceq n^{-1/r}, & \text{for cyclic } R \end{cases} \tag{3.16}$$

It turns out that the Berry-Esseen bound in this setting would be slower than $n^{-1/2}$, unlike that in i.i.d. and noiseless U-statistic settings. The exact reason is technical and will be better seen in the proof of Theorem 3.4, but the intuitive explanation is that if ρ_n is too small, the higher degree ($\geqslant 2$) random errors in $\widehat{U}_n - U_n$ diminishes too slowly compared to the scale of the demoninator of $\widehat{T}_n$. If the network sparsity ρ_n falls below the typically assumed lower bounds: n^{-1} for acyclic R and $n^{-2/r}$ for cyclic R [20, 17, 61], no known consistency guarantee exists. In fact, in this case we do not even know if $\widehat{T}_n$ is asymptotically normal.

THEOREM 3.4. *Under the conditions of Lemma 3.1, except replacing Condition (ii) by (3.16), we have the following modified Berry-Esseen bound*

$$\left\| F_{\widehat{T}_n}(u) - G_n(u) \right\|_\infty \asymp \left\| F_{\widehat{T}_n}(u) - \Phi(u) \right\|_\infty = O\left(\mathcal{M}(\rho_n, n; R)\right) \bigwedge o(1),$$

where recall that $\Phi(\cdot)$ is the CDF of $N(0,1)$. Moreover,

$$\left\| F_{\widehat{T}_n}(u) - \widehat{G}_n(u) \right\|_\infty = \widetilde{O}_p\left(\mathcal{M}(\rho_n, n; R)\right) \bigwedge o_p(1)$$

In the sparse regime, the current upper bound on the remainder terms would dominate the $n^{-1/2}$ leading term in the Edgeworth expansion. In other words, in sparse networks, the Edgeworth expansion is guaranteed by the same error rate bound as a simple $N(0,1)$ approximation. On the other hand, the conclusion of Theorem 3.4 connects the error bound results for dense and sparse regimes. Interestingly, as the order of ρ_n decreases from $n^{-1/2}$ to n^{-1} for acyclic R, or from $n^{-1/r}$ to $n^{-2/r}$ for cyclic R, we see a gradual depreciation in the uniform CDF approximation error from the order of $n^{-1/2}$ to merely uniform consistency. The classical literature only studied the boundary cases ($\rho_n = \omega(n^{-1})$ or $\rho_n = \omega(n^{-2/r})$, depending on R), and our result here reveals the complete picture.

A natural question is whether a higher-order approximation would be possible in the sparse regime. We conjecture not. We also conjecture that the Berry-Esseen bound that both empirical Edgeworth expansion and $N(0,1)$ approximation achieve is either sharp or nearly sharp, but we do not know an answer for sure. This would be an interesting future work.

3.6. *Comparison table of our method to benchmarks.* We conclude this section by comparing our results to some representative works in classical and very recent literature.

TABLE 2
Comparison of CDF approximation methods for noisy/noiseless studentized U-statistics

Method	U-stat. type	Popul. momt.[17]	Smooth graphon	Non lat. /Cramer	Network sparsity assumption on ρ_n[18]	CDF approx. error rate
Our method (empirical Edgeworth)	Noisy	No	No	If yes	$\omega(n^{-2/r})$(C); $\omega(n^{-1})$(Ac)[19]	$\widetilde{O}_p\left(\mathcal{M}(\rho_n, n; R)\right) \wedge o_p(1)$ **(H)**[20]
				If no	$\omega(n^{-2/r})$(C); $\omega(n^{-1})$(Ac) and $O\left((\log n)^{-1}\right)$(C, Ac)	$\widetilde{O}_p\left(\mathcal{M}(\rho_n, n; R)\right) \wedge o_p(1)$ **(H)**
Node re-/sub- sampling justified by our theory	Noisy	No	No	Yes	$\omega(n^{-1/r})$(C); $\omega(n^{-1/2})$(Ac)	$o_p(n^{-1/2})$ **(H)**
Bickel, Chen and Levina [20]	Noisy	No[21]	No	No	$\omega(n^{-2/r})$(C); $\omega(n^{-1})$(Ac)	Consistency
Bhattacharyya and Bickel [17]	Noisy	No	No	No	$\omega(n^{-2/r})$(C); $\omega(n^{-1})$(Ac)	Consistency
Green and Shalizi [61]	Noisy	No	Mixed[22]	No	R is Ac; or $\omega(n^{-1/(2r)})$(C)[23]	Consistency
Levin and Levina [93]	Noisy	No	Low-rank[24]	No	$\omega(n^{-1} \cdot \log n)$ (Ac*)[25]	Consistency
Bickel, Gotze and Zwet [21]	Noiseless	Yes	No	Yes	Not applicable	$o(n^{-1})$ **(H)**
Bentkus, Gotze and Zwet [14]	Noiseless	Yes	No	Yes	Not applicable	$O(n^{-1})$ **(H)**
Putter and Zwet [110]	Noiseless	No	No	Yes	Not applicable	$o_p(n^{-1/2})$ **(H)**
Bloznelis [23]	Noiseless	No	No	Yes	Not applicable	$o_p(n^{-1/2})$ **(H)**

4. Theoretical and methodological applications.

4.1. *Higher-order accuracy of node sub- and re-sampling network bootstraps.* One important corollary of our results is first higher-order accuracy proof of some network bootstrap schemes. For a network bootstrap scheme that produces an estimated $\widehat{U}_{n*}^{b}$ and its jackknife[26] variance estimator $\widehat{S}_{n*}^{*}$, define $\widehat{T}_{n*}^{*} = (\widehat{U}_{n*}^{b} - \widehat{U}_n)/\widehat{S}_{n*}^{*}$. We are going to establish the first explicit rate guarantees for following two schemes.

(a). Sub-sampling [17]: randomly sample n^* nodes from $\{1, \ldots, n\}$ *without replacement*, and compute $\widehat{T}_{n*}^{*}$ from the induced sub-network of A.
(b). Re-sampling [61]: random sample n nodes from $\{1, \ldots, n\}$ *with replacement*, and compute $\widehat{T}_{n*}^{*}$ from the induced sub-network of A.

[17] "Yes" means need to know the population moments that appear in Edgeworth coefficients, i.e. ξ_1, $\mathbb{E}[g_1^3(X_1)]$ and $\mathbb{E}[g_1(X_1)g_1(X_2)g_2(X_1, X_2)]$.

[18] To compare ρ_n assumptions, see our Remark 3.2

[19] (C): **c**yclic R; (Ac): **ac**yclic R.

[20] Recall $\mathcal{M}(\rho_n, n; R)$ defined in (1.1) and $\widetilde{O}_p$ defined in Section 1.4. **(H)**: higher-order accuracy results. "Consistency": only convergence, no error rate.

[21] In [20, 17], $\widehat{U}_n - \mu_n$ was rescaled by ρ_n and n. Whether assuming the knowledge of the true ρ_n or not does not matter for their $o_p(1)$ error bound, but it would make a difference if an $o_p(n^{-1/2})$ or finer bound is desired.

[22] The bootstrap based on denoised A requires smoothness. See Theorem 2 of [61].

[23] It seems their assumption for cyclic R was a typo, and $\rho_n = \omega(n^{-2/r})$ should suffice. Also, they used [17] in their proof, which requires $\rho_n = \omega(n^{-1})$ for (Ac).

[24] [93] assumed the graphon rank is low and known.

[25] (Ac*): They require the motif to be either acyclic or an r-cycle, see their Theorem 4. Their Theorem 3 requires condition (8) that only holds when R is a clique.

[26] Here, we use jackknife variance estimator in bootstraps to better connect with existing literature in the proof.

REMARK 4.1. *Notice that [61] did not study the studentized form, and [17] propoosed a different variance estimator (what they call "$\widehat{\sigma}_{B_i}$"). Our justifications focus on the* sampling schemes *combined with some natural formulation, not necessarily the same formulation as in these two papers.*

REMARK 4.2. *As noted in [61], scheme (b) can be viewed as our data generation procedure described in Sections 2.1 and 2.2 but with the graphon f replaced by the adjacency-induced graphon $A(u,v) = A_{\lceil nu \rceil, \lceil nv \rceil}$, where $\lceil y \rceil := \text{Ceiling}(y)$. This may seem similar to f-based data generaion, but in fact they are distinct. The graphon $A(\cdot,\cdot)$ inherits the binary nature of A and will necessarily yield a lattice $g_1^*(X_1^*)$ regardless of the original graphon f and the motif $\mathcal{R}$, rendering most classical Edgeworth analysis techniques inapplicable. But the real obstacle is that the bootstrapped network data from $A(\cdot,\cdot)$ have no edge-wise observational errors (i.e. no counterpart to the randomness in $A|W$). Consequently, $\widehat{T}^*_{n*}$ loses the self-smoothing feature that $\widehat{T}_n$ enjoys. For this reason, when justifying the higher-order accuracy of network bootstraps, we cannot simply reproduce the proof of our main theorem that crucially benefits from the self-smoothing effect. Aligned with this observation, the eventual error rates that we established for network bootstraps are significantly worse than our population and empirical Edgeworth expansions. We conjecture that further improving the error guarantee for network bootstraps beyond Theorem 4.1, if possible, might require much more sophisticated analysis.*

THEOREM 4.1. *Assume $g_1(X_1)$ satisfies a Cramer's condition such that $\limsup_{t\to\infty} \left| \mathbb{E}\left[e^{\mathrm{i}tg_1(X_1)\cdot\xi_1^{-1}} \right] \right| < 1$. Under the conditions of Theorem 3.2, we conclude for the following bootstrap schemes:*

(a). *Sub-sampling: choosing $n^* \asymp n$ and $n - n^* \asymp n$, we have*

$$\left\| F_{\widehat{T}^*_{n^*}}(u) - F_{\widehat{T}_{n^*(1-n^*/n)}}(u) \right\|_\infty = o_p\left((n^*)^{-1/2} \right) = o_p(n^{-1/2}). \tag{4.1}$$

(b). *Re-sampling: choosing $n^* = n$, we have*

$$\left\| F_{\widehat{T}^*_{n^*}}(u) - F_{\widehat{T}_{n^*}}(u) \right\|_\infty = o_p\left((n^*)^{-1/2} \right) = o_p(n^{-1/2}). \tag{4.2}$$

REMARK 4.3. *In the proof of Theorem 4.1, we combined our main results with the results of [23] for finite population U-statistics. It is important to notice that all existing works under the finite populations did assume non-lattice with population size growing to infinity, see condition (1.13) in Theorem 1 of [23]. Consequently, the higher-order accuracy of some network bootstraps is only proved under Cramer's condition by so far.*

Part (a) of Theorem 4.1 quantifies the *effective sample size* in the sub-sampling network bootstrap: sampling n^* out of n nodes without replacement, the resulting bootstrap $\widehat{T}^*_{n^*}$ approximates the distribution of $\widehat{T}_m$ where $m = \{n^*/n \cdot (1 - n^*/n)\} \times n$. Consequently, in order to approach the sampling distribution of $\widehat{T}_n$ with higher-order accuracy using sub-sampling [17], one must have an observed network of at least $4n$ nodes, from which she shall repeatedly sub-sample $2n$ nodes without replacement.

4.2. *One-sample t-test for network moments under general null graphon models.* In this and the next subsections, we showcase how our results immediately lead to useful inference procedures for network moments. For a given motif R, we test on its population mean frequency μ_n. Since μ_n depends on n through ρ_n, we formulate the hypotheses as follows

$$H_0 : \mu_n = c_n, \text{ versus } H_a : \mu_n \neq c_n.$$

where c_n is a speculated value of $\mu_n = \mathbb{E}[h(A_{1,\dots,r})]$. In practice, c_n may come from a prior study on a similar data set or fitting a speculated model to the data (for concrete examples on c_n guesses, see Section 6.1 of [17]).

Here for simplicity we only discuss a two-sided alternative, and one-sided cases are exactly similar. The p-value can be formulated using our empirical Edgeworth expansion $\widehat{G}_n(\cdot)$ in (3.14):

$$\text{Estimated p-value} = 2 \cdot \min\left\{\widehat{G}_n(t^{(\text{obs})}), 1 - \widehat{G}_n(t^{(\text{obs})})\right\}. \tag{4.3}$$

where $t^{(\text{obs})} := (\widehat{u}_n^{(\text{obs})} - c_n)/\widehat{s}_n^{(\text{obs})}$, and $\widehat{u}_n^{(\text{obs})}$ and $\widehat{s}_n^{(\text{obs})}$ are the observed $\widehat{U}_n$ and $\widehat{S}_n$, respectively. We have the following explicit Type-II error rate.

THEOREM 4.2. *Under the conditions of Theorem 3.2, we have the following results:*

1. *The Type-I error rate of test* (4.3) *is* $\alpha + O(\mathcal{M}(\rho_n, n; R))$.
2. *The Type-II error rate of this test is* $o(1)$ *when* $|c_n - \mu_n| = \omega\left(\rho_n^s \cdot n^{-1/2}\right)$.

REMARK 4.4. *The null model we study is complementary to the degenerate Erdos-Renyi null model in [91, 54, 55]. The scientific questions are also different: they test model goodness-of-fit whereas we test population moment values. Notice that distinct network models may possibly share some common population moments. These approaches also use very different methods and analysis techniques.*

4.3. *Cornish-Fisher confidence intervals for network moments.* Noticing that $\widehat{G}_n$ is almost never a valid CDF, in order to preserve the higher-order accuracy of $\widehat{G}_n$, we use the Cornish-Fisher expansion [44, 53] to approximate the quantiles of $F_{\widehat{T}_n}$. A Cornish-Fisher expansion is the inversion of an Edgeworth expansion, and its validity hinges on the validity of its corresponding Edgeworth expansion. Using the technique of [65], we have

THEOREM 4.3. *For any* $\alpha \in (0,1)$, *define the lower* α *quantile of the distribution of* $\widehat{T}_n$ *by*

$$q_{\widehat{T}_n;\alpha} := \arg\inf_{q\in\mathbb{R}} F_{\widehat{T}_n}(q) \geqslant \alpha \tag{4.4}$$

and define the approximation

$$\widehat{q}_{\widehat{T}_n;\alpha} := z_\alpha - \frac{1}{\sqrt{n}\cdot\widehat{\xi}_1^3}\cdot\left\{\frac{2z_\alpha^2+1}{6}\cdot\widehat{\mathbb{E}}[g_1^3(X_1)]\right.$$

$$\left. + \frac{r-1}{2}\cdot\left(z_\alpha^2+1\right)\widehat{\mathbb{E}}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)]\right\}, \tag{4.5}$$

where $z_\alpha := \Phi^{-1}(\alpha)$. *Then under the conditions of Theorem 3.2, we have*

(a). *The discrepancy between nominal and actual percentage-below for* $q_{\widehat{T}_n;\alpha}$ *is bounded by*

$$|F_{\widehat{T}_n}(q_{\widehat{T}_n;\alpha}) - \alpha| = O(\mathcal{M}(\rho_n, n; R)) \tag{4.6}$$

(b). *The "horizontal" error bound:*

$$\left|\widehat{q}_{\widehat{T}_n;\alpha} - q_{\widehat{T}_n;\alpha}\right| = \widetilde{O}_p(\mathcal{M}(\rho_n, n; R)) \tag{4.7}$$

(c). *The uniform "vertical" error bound:*

$$\mathbb{P}(\widehat{T}_n \leqslant \widehat{q}_{\widehat{T}_n;\alpha}) = \alpha + O(\mathcal{M}(\rho_n, n; R)). \tag{4.8}$$

The vertical error bound describes the approximation error between the nominal and actual coverage probabilities, whereas the horizontal error bound governs the approximation of quantiles. Using the vertical error bound, a $1-\alpha$ two-sided symmetric Cornish-Fisher confidence interval for estimating μ_n can be easily constructed as follows

$$\left(\widehat{U}_n - \widehat{q}_{\widehat{T}_n;1-\alpha/2}\cdot\widehat{S}_n, \widehat{U}_n - \widehat{q}_{\widehat{T}_n;\alpha/2}\cdot\widehat{S}_n\right) \tag{4.9}$$

and by Theorem 4.3, we know this CI has a $1-\alpha+O\left(\mathcal{M}(\rho_n, n; R)\right)$ coverage probability. One-sided confidence intervals can be constructed exactly similarly, thus we omit them.

5. Simulations.

5.1. *Simulation 1: Higher-order accuracy of empirical Edgeworth expansion.* In the first simulation, our numerical studies focus on the CDF of $F_{\widehat{T}_n}$. In an illustrative example, we simulate with a lattice $g_1(X_1)$ and show the distinction between $F_{\widehat{T}_n}$ and F_{T_n} that clearly illustrates the self-smoothing effect in $\widehat{T}_n$. Then we systematically compare the performance of our empirical Edgeworth expansion to benchmarks that demonstrates the clear advantage of our method in both accuracy and computational efficiency.

We begin by describing the basic settings. We range the network size n in an exponentially spaced set $n \in \{10, 20, 40, 80, 160\}$. Synthetic network data are generated from three graphons marked by their code-names in our figures: (1). `"BlockModel"`: This is an ordinary stochastic block model with $K = 2$ equal-sized communities and the following edge probabilities $B = (0.6, 0.2; 0.2, 0.2)$; (2). `"SmoothGraphon"`: Graphon 4 in [136], i.e. $f(u,v) := (u^2+v^2)/3\cdot\cos(1/(u^2+v^2))+0.15$. This graphon is smooth and full-rank [136]; (3). `"NonSmoothGraphon"`[40]: We set up a high-fluctuation area in a smooth f to emulate the sampling behavior of a non-smooth graphon, as follows

$$f(u,v) := 0.5\cos\left\{0.1/((u-1/2)^2+(v-1/2)^2)^{-1}+0.01\right\}\max\{u,v\}^{2/3}+0.4.$$

We test the four simplest motifs: *edge*, *triangle*, *V-shape*[27], and a *three-star* among 4 nodes with edge set $\{(1,2),(1,3),(1,4)\}$. The main computation bottleneck lies in the evaluation of $F_{\widehat{T}_n}$. Network bootstraps also becomes costly as n increases.

The benchmarks are: 1. $N(0,1)$ (its computation time is deemed zero and not compared to others); 2. sub-sampling by [17] with $n^* = n/2$; 3. re-sampling A by [61]; 4. latent space bootstrap called "ASE plug-in" defined in Theorem 2 of [93]. Notice that we equipped [93] with an adaptive network rank estimation[28] by USVT [35].

For each (graphon, motif, n) tuple, we first evaluate the true sampling distribution of $\widehat{T}_n$ by a Monte-Carlo approximation that samples $n_{\rm MC} := 10^6$ networks from the graphon. Next we start 30 repeated experiments: in each iteration, we sample A from the graphon and approximate $F_{\widehat{T}_n}$ by all methods, in which we draw $n_{\rm boot} = 2000$ bootstrap samples for each bootstrap method – notice that this is 10 times that in [93]. We compare

$$\mathrm{Error}(\widehat{F}_{\widehat{T}_n}) := \sup_{u\in[-2,2];10u\in\mathbb{Z}}\left|\widehat{F}_{\widehat{T}_n}(u) - F_{\widehat{T}_n}(u)\right|. \tag{5.1}$$

REMARK 5.1. *We need many Monte-Carlo repetitions, because the uniform accuracy of the empirical CDF of an i.i.d. sample is only $O_p(n_{\rm MC}^{-1/2})$ [50, 88], and for the noiseless and*

[27] A "V-shape" is the motif obtained by disconnecting one edge in a triangle. In the language of [20], it is a 2-star.

[28] Consequently, our enhanced version of this benchmark can decently denoise some smooth but high-rank graphons, in view of the remarks in [136] and the results of [134].

noisy U-statistic setting, the bound might be worse than the i.i.d. setting due to dependency[29]. *Therefore, we set* $n_{\rm MC} \gg \max(n^2) = 160^2$ *to prevent the errors of the compared methods being dominated by the error of the Monte-Carlo procedure; while keep our simulations reproducible with moderate computation cost. We did find smaller* $n_{\rm MC}$ *such as* 10^5 *to cloud the performance of our method.*

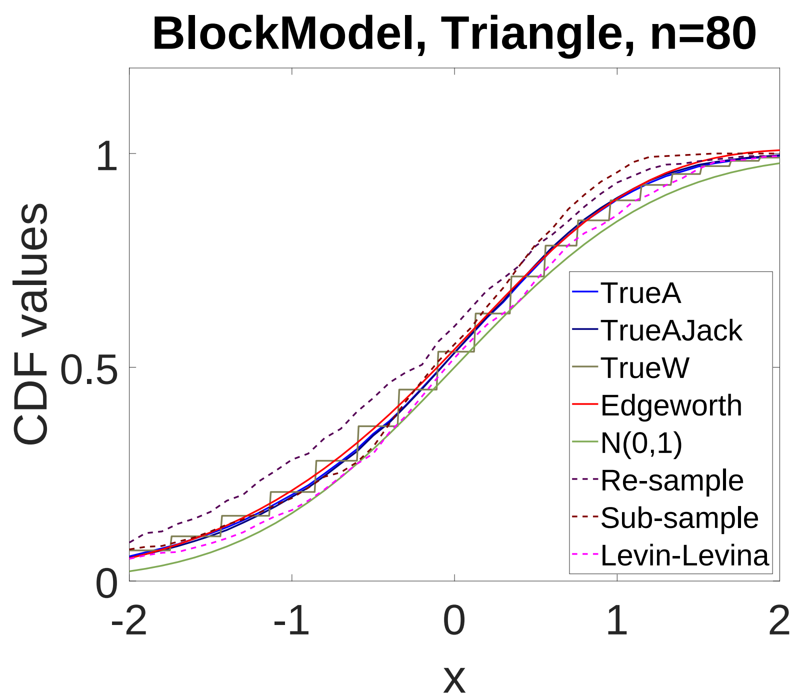

Fig 1: CDF curves of the studentization forms and approximations. Network size $n = 80$. The graphon is the "`BlockModel`" we described earlier in this section, and the motif is triangular. Each bootstrap method draws 500 random samples. `TrueA` is $F_{\widehat{T}_n}$; `TrueAJack` is $F_{\widehat{T}_{n;\rm jackknife}}$; `TrueW` is F_{T_n}; `Edgeworth` is our empirical Edgeworth expansion; `Re-sample` is node re-sampling A in [61]; `Sub-sample` is node sub-sampling A in [17]; `Levin-Levina` is the "ASE plug-in" bootstrap in [93].

Now we present the results. We first present the illustrative simulation for just one specific setting. Figure 1 shows the distribution approximation curves under a block model graphon that yields a lattice $g_1(X_1)$. Lines correspond to the population CDF of $\widehat{T}_n$, its jackknife version and noiseless version, all evaluated by Monte-Carlo procedures; our proposed empirical Edgeworth expansion; and benchmarks. We make two main observations. First, `TrueA` and `TrueAJack` are almost indistinguishable, echoing our Theorem 3.3; meanwhile, they are both smooth and rather different from the step-function `TrueW`. This clearly demonstrates the self-smoothing feature of $\widehat{T}_n$ in the lattice case. If we change the graphon to a smooth one, these curves would all be smooth and close to each other. Second, we observe the higher accuracy of our empirical Edgeworth expansion compared to competing methods. In fact,

[29]This is not to be confused with the Edgeworth approximation error bound. In this Monte Carlo procedure, both the true and approximate $F_{\widehat{T}_n}$ are oracle.

repeating this experiment multiple times, our method shows significantly stabler approximations than bootstraps.

Next, we conduct a systematic comparison of the performances of all methods across many settings. We mainly varied three factors: graphon type, motif type and network size, over the previously described ranges. Our experiment results under different network sparsity levels would have to sink to Supplemental Material due to page limit, and here we keep $\rho_n = 1$. Results are shown in Figure 2 (error) and Figure 3 (time cost), where error bars show standard deviations.

In all experiments, our empirical Edgeworth expansion approach exhibited clear advantages over benchmark methods in all aspects: the absolute values of errors, the diminishing rates of errors, and computational efficiency. Our method shows a higher-order accuracy by slopes steeper than $-1/2$ and much steeper than other methods. On computation efficiency, our method is the second cheapest after the simple $N(0,1)$ approximation (that does not need computation) and much faster than network bootstraps. It typically costs about $e^{-5} \approx 1/150$ the time of sub-sampling and about $e^{-7} \approx 1/1000$ the time of re-sampling. Our method only needs one run and does not require repeated sampling.

Notice that there is no simple rule to judge the difficulty of different scenarios, which jointly depends on the graphon and the motif through implicit and complex relationship. In our experience, triangle may be more difficult than V-shape under some graphons, but easier under some others, and this comparison may vary from method to method. Answering this question requires calculation of the population Edgeworth expansion up to $o(n^{-1})$ remainder, and the leading term in the remainder of the one-term Edgeworth expansion would then quantify the real difficulty. But the calculation is very complicated and outside the scope of this paper.

We did not observe the higher-order accuracy of bootstrap methods as our results predicted. One likely reason is the numerical accuracy limited by the n_{boot} that our computing servers can afford. We did see an observable improvement in the performances of network bootstraps as we increased n_{boot} from 200 suggested by [93] to the current 2000. But further increasing n_{boot} will also increase their time costs and potentially memory usage. We ran each experiment on 36 parallel Intel(R) Xeon(R) X5650 CPU cores at 2.67GHz with 12M cache and 2GB RAM. It took roughly 3∼8 hours to run each experiment that produces one individual plot in Figures 2 and 3.

5.2. *Simulation 2: Finite-sample performance of Cornish-Fisher confidence interval.* In this simulation, we numerically assess the performance of our Cornish-Fisher confidence interval, compared to benchmark methods. Throughout this subsection, we set $\alpha = 0.2$ and focus on symmetric two-sided confidence intervals. We inherit most simulation settings from Section 5.1 with some modifications we now clarify. The main difference is that in this simulation, we must conduct many repeated experiments in order to accurately evaluate the coverage probability (each iteration produces a binary outcome of whether the CI contains the population parameter). We repeated the experiment 10000 times for our method and normal approximation, and 500 times for the much slower bootstrap methods. Due to the computer limitations, while we can keep the same number of Monte Carlo evaluations, in order to repeat the entire experiment 500 times to accurately evaluate the actual CI coverage rates of bootstraps, we have to reduce their numbers of bootstrap samples to 500 (still more than the 200 in [93]). We evaluate three performance measures: `coverage`: *actual coverage probability*; `length`: *confidence intereval length*; and `time`: *computation time in seconds*.

Due to page limit, in the main text, we only present the results for the setting $n = 80$ and $\rho_n = 1$ in Tables 3 (block model), 4 (smooth graphon) and 5 (non-smooth graphon). Each entry is formatted "mean(standard deviation)". We sink the remaining results to Supplemental

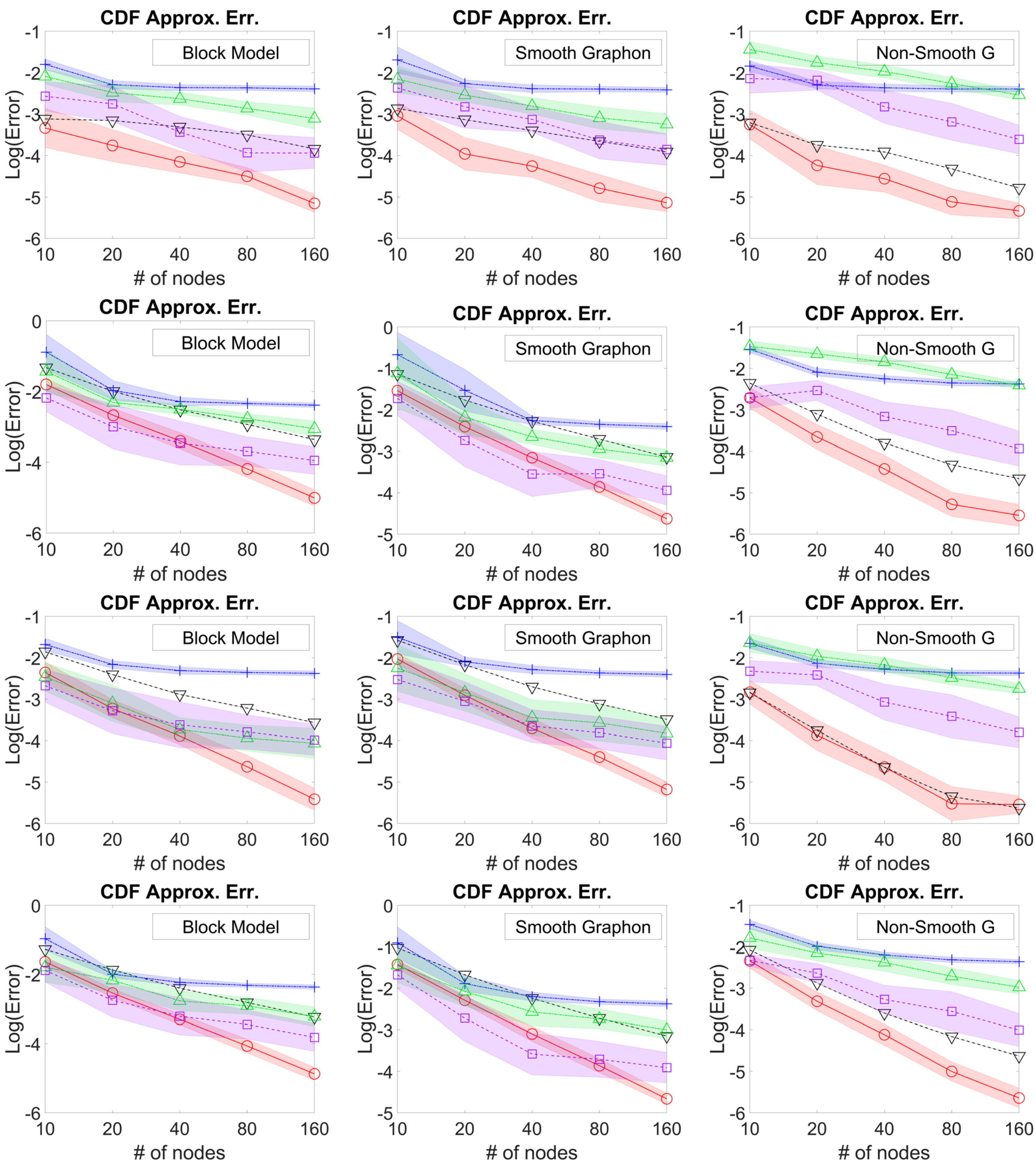

Fig 2: CDF approximation errors. Both axes are log(e)-scaled. **Motifs:** row 1: `Edge`; row 2: `Triangle`; row 3: `Vshape`; row 4: `ThreeStar`. Red solid curve marked circle: our method (empirical Edgeworth); black dashed curve marked down-triangle: $N(0,1)$ approximation; green dashed curve marked up-triangle: re-sampling of A in [61]; blue dashed curve marked plus: [17] sub-sampling $\asymp n$ nodes; magenta dashed line with square markers: ASE plug-in bootstrap in [93].

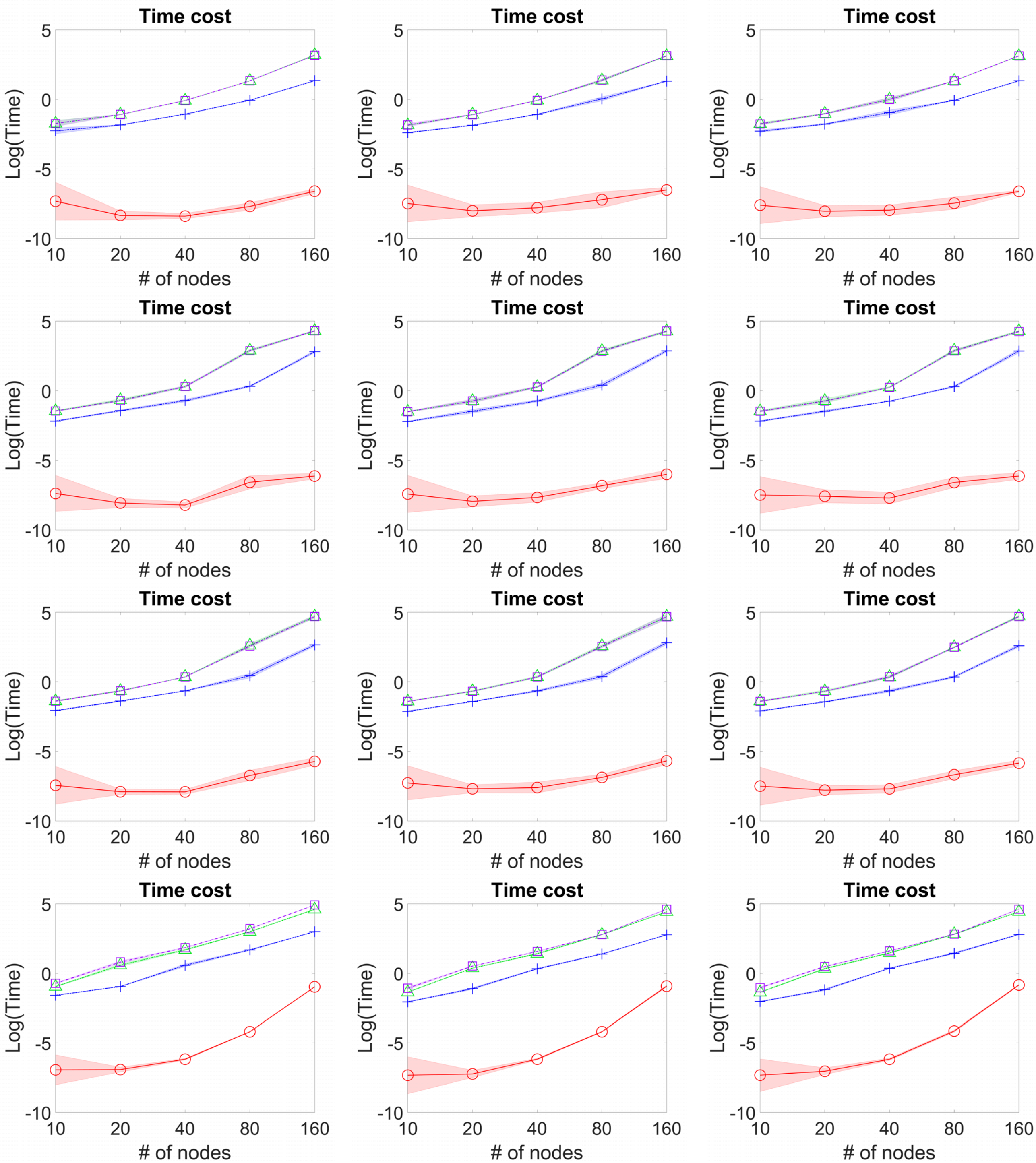

Fig 3: Time costs (in seconds) of all methods. Both axes are log(e)-scaled. **Motifs:** row 1: `Edge`; row 2: `Triangle`; row 3: `Vshape`; row 4: `ThreeStar`. Red solid curve marked circle: our method (empirical Edgeworth); green dashed curve marked up-triangle: re-sampling of A in [61]; blue dashed curve marked plus: [17] sub-sampling $\asymp n$ nodes; magenta dashed line with square markers: ASE plug-in bootstrap in [93]. We regarded $N(0,1)$ as zero time cost so does not appear in the time cost plot.

TABLE 3
Performance measures of 95% confidence intervals
$n = 80$, $\rho_n \asymp 1$, *graphon: block model*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.957(0.202)$	$0.953(0.211)$	$0.956(0.205)$	$0.952(0.213)$
	Length $= 0.097(0.010)$	$0.040(0.008)$	$0.200(0.033)$	$0.145(0.033)$
	LogTime $= -8.448(0.110)$	$-7.214(0.083)$	$-7.165(0.082)$	$-7.180(0.353)$
Norm. Approx.	$0.950(0.218)$	$0.934(0.248)$	$0.942(0.235)$	$0.932(0.251)$
	$0.097(0.010)$	$0.040(0.008)$	$0.200(0.033)$	$0.145(0.033)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.842(0.365)$	$0.870(0.337)$	$0.852(0.355)$	$0.852(0.355)$
	$0.068(0.009)$	$0.031(0.007)$	$0.147(0.026)$	$0.113(0.025)$
	$-2.591(0.008)$	$-2.160(0.026)$	$-2.127(0.024)$	$-0.992(0.006)$
Green and Shalizi [61]	$0.938(0.241)$	$0.944(0.230)$	$0.934(0.249)$	$0.938(0.241)$
	$0.096(0.013)$	$0.044(0.010)$	$0.204(0.038)$	$0.150(0.037)$
	$-1.198(0.007)$	$0.499(0.032)$	$0.142(0.035)$	$0.383(0.010)$
Levin and Levina [93]	$0.942(0.234)$	$0.942(0.234)$	$0.942(0.234)$	$0.942(0.234)$
	$0.099(0.013)$	$0.043(0.010)$	$0.209(0.039)$	$0.155(0.038)$
	$-1.188(0.004)$	$0.507(0.028)$	$0.142(0.027)$	$0.489(0.004)$

TABLE 4
Performance measures of 95% confidence intervals
$n = 80$, $\rho_n \asymp 1$, *graphon: smooth graphon*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.958(0.201)$	$0.940(0.238)$	$0.951(0.216)$	$0.942(0.235)$
	Length $= 0.092(0.009)$	$0.021(0.005)$	$0.141(0.025)$	$0.083(0.021)$
	LogTime $= -8.225(0.113)$	$-7.363(0.066)$	$-7.278(0.086)$	$-6.974(0.541)$
Norm. Approx.	$0.951(0.216)$	$0.920(0.271)$	$0.938(0.242)$	$0.923(0.266)$
	$0.092(0.009)$	$0.021(0.005)$	$0.141(0.025)$	$0.083(0.021)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.816(0.388)$	$0.840(0.367)$	$0.826(0.379)$	$0.852(0.355)$
	$0.066(0.009)$	$0.018(0.005)$	$0.110(0.021)$	$0.072(0.018)$
	$-2.554(0.010)$	$-2.124(0.026)$	$-2.139(0.026)$	$-1.020(0.027)$
Green and Shalizi [61]	$0.928(0.259)$	$0.946(0.226)$	$0.938(0.241)$	$0.948(0.222)$
	$0.092(0.012)$	$0.025(0.007)$	$0.147(0.029)$	$0.090(0.024)$
	$-1.144(0.009)$	$0.497(0.042)$	$0.157(0.054)$	$0.334(0.025)$
Levin and Levina [93]	$0.948(0.222)$	$0.948(0.222)$	$0.950(0.218)$	$0.958(0.201)$
	$0.095(0.012)$	$0.024(0.007)$	$0.153(0.030)$	$0.095(0.025)$
	$-1.138(0.005)$	$0.507(0.031)$	$0.172(0.030)$	$0.447(0.019)$

Materials. Our method exhibits very accurate actual coverage probabilities consistently close to the nominal confidence level. Our method is the only method that can always achieve a $\leqslant 0.010$ coverage error across all settings. It also produces competitively short confidence interval lengths, again, reflecting the high accuracy of the method. The comparison of computational efficiency between different methods echoes the qualitative results in Figure 3 despite slightly different settings and confirms our method's huge speed advantage over all bootstrap methods.

It is interesting to observe that under the setting of this simulation, our empirical Edgeworth expansion method always produces the same interval length as the normal approximation. This is not a coincidence in view of (4.5), (4.9) and that $z^2_{\alpha/2} = z^2_{1-\alpha/2}$. In other words, as long as the studentization form $\widehat{T}_n$ that $N(0,1)$ approximates is equipped with the same variance estimator $\widehat{S}_n$ as our method, our two-sided Edgeworth confidence interval is a bias-corrected version (by mean-shift) of the corresponding ordinary CLT confidence interval.

TABLE 5
Performance measures of 95% confidence intervals
$n = 80$, $\rho_n \asymp 1$, *graphon: non-smooth graphon*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.956(0.205)$	$0.957(0.203)$	$0.957(0.202)$	$0.957(0.203)$
	Length $= 0.116(0.009)$	$0.135(0.010)$	$0.422(0.027)$	$0.531(0.040)$
	LogTime $= -8.291(0.076)$	$-7.345(0.103)$	$-7.817(0.153)$	$-7.045(0.373)$
Norm. Approx.	$0.952(0.215)$	$0.949(0.220)$	$0.951(0.215)$	$0.950(0.218)$
	$0.116(0.009)$	$0.135(0.010)$	$0.422(0.027)$	$0.531(0.040)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.830(0.376)$	$0.832(0.374)$	$0.830(0.376)$	$0.836(0.371)$
	$0.081(0.010)$	$0.096(0.011)$	$0.297(0.031)$	$0.379(0.041)$
	$-2.569(0.012)$	$-2.105(0.051)$	$-2.116(0.035)$	$-1.011(0.005)$
Green and Shalizi [61]	$0.940(0.238)$	$0.938(0.241)$	$0.944(0.230)$	$0.944(0.230)$
	$0.112(0.013)$	$0.135(0.014)$	$0.415(0.041)$	$0.529(0.055)$
	$-1.201(0.011)$	$0.547(0.075)$	$0.169(0.037)$	$0.328(0.015)$
Levin and Levina [93]	$0.954(0.210)$	$0.956(0.205)$	$0.956(0.205)$	$0.954(0.210)$
	$0.116(0.013)$	$0.138(0.013)$	$0.427(0.039)$	$0.544(0.052)$
	$-1.190(0.003)$	$0.534(0.049)$	$0.162(0.033)$	$0.436(0.014)$

In the Supplemental Materials, we present simulation results for the remaining configurations among $n \in \{80, 160\}$ and $\rho_n \asymp \{1, n^{-1/4}, n^{-1/2}, n^{-1}\}$. For very sparse networks, our method and $N(0,1)$ approximation produce similar conservative confidence intervals for the $R =$ Edge. On the other hand, all methods fail spectacularly for more complex motifs. Despite the required ρ_n lower bounds for all acyclic motifs are identically $\omega(n^{-1})$ for our method and $N(0,1)$ approximation, the results are not surprising for two reasons: (i) the theory requires $\rho_n \gg n^{-1}$, so the simulation setting $\rho_n \asymp n^{-1}$ is the boundary case and sensible outputs are not guaranteed; and (ii) the constant factor may matter a lot, and different acyclic motif shapes may require different minimum constants factor in ρ_n to show sensible results.

5.3. *Simulation 3: Numerical evaluation of the finite-sample impact of sparsity.* In this part, we conduct numerical studies to evaluate the finite sample performances of our method compared to benchmarks as the network grows sparser under fixed n. Despite in Simulation 5.1, we tested different network sparsity settings (see Supplemental Material), it would still be interesting to more directly illustrate the impact of ρ_n for each fixed network size. The simulation set up carries over the same set of graphon models, motif shapes and compared methods from Simulation 5.1. Here, for simplicity, we only tested $n = 80, 160$ and varied ρ_n in a wider range of sparsity as follows: $\{1$ ("dense"), $n^{-1/4}, n^{-1/2}, n^{-1}\}$.

Figure 4 shows the CDF approximation errors under different ρ_n settings for $n = 160$. Aligned with our theory's prediction, we observe that as the network grows sparser, our method's performance depreciates and gradually regresses to the performance of normal approximation. Due to page limit, we sink the approximate error plots for $n = 80$ and the time cost plots for both n settings to Supplemental Materials.

5.4. *Simulation 4: Degree-corrected stochastic block models.* We also tested our method on networks with degree heterogeneity. Our method maintains significant advantage in both accuracy and speed. Due to page limit, we sink all results and interpretation to Supplemental Materials (See Section 9.4).

5.5. *Simulation 5: Scalability of our method to large networks.* In this experiment, we test the scalability of our method. We find that all the three benchmark methods that we tested in previous simulations would fail to finish running on our high performance computing servers within 24 hours. Therefore, only the time costs of our method are recorded.

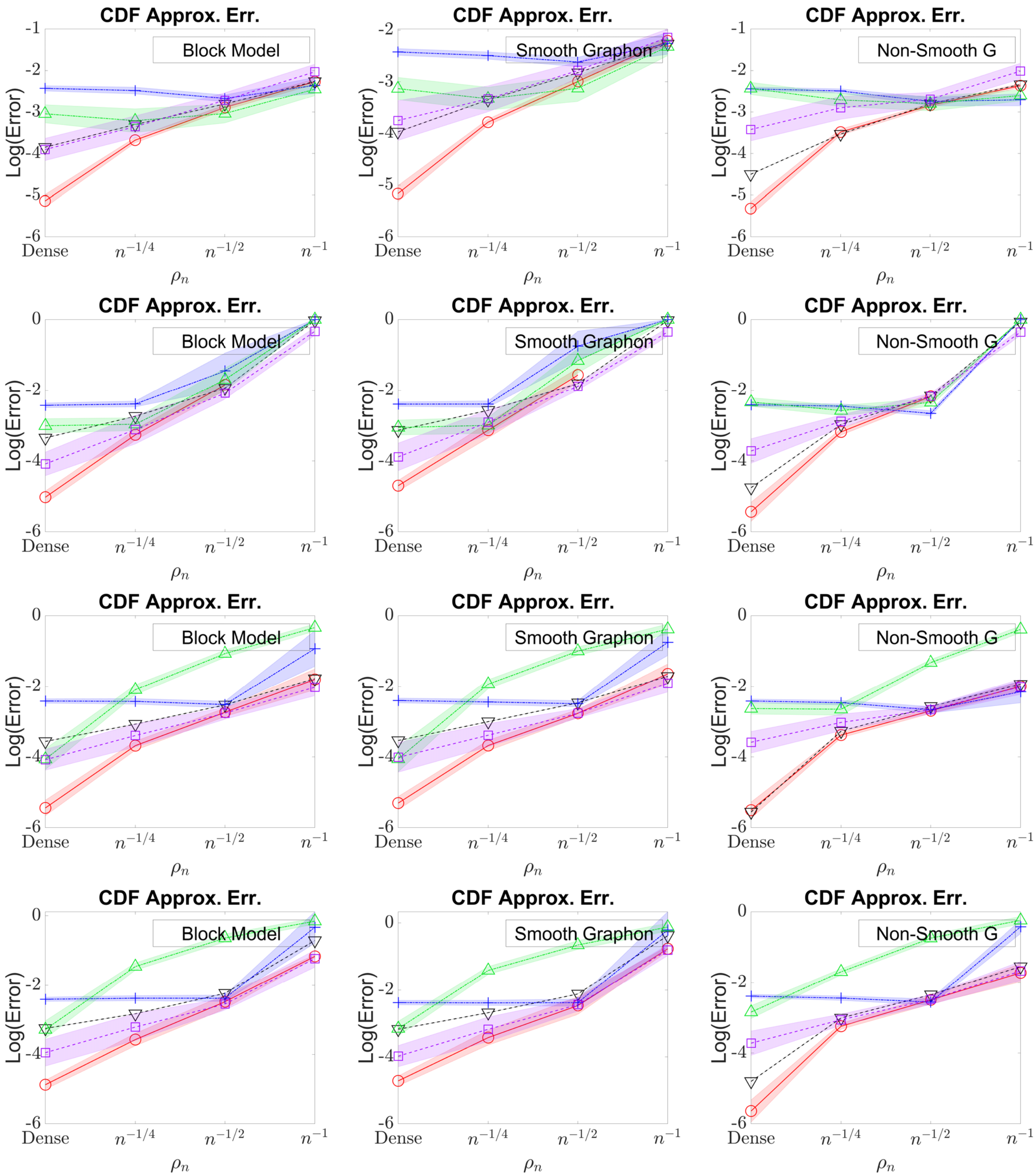

Fig 4: Impact of sparsity on approximation errors, $n = 160$. Both axes are log(e)-scaled. **Motifs:** row 1: `Edge`; row 2: `Triangle`; row 3: `Vshape`; row 4: `ThreeStar`. Red solid curve marked circle: our method (empirical Edgeworth); black dashed curve marked down-triangle: $N(0, 1)$ approximation; green dashed curve marked up-triangle: re-sampling of A in [61]; blue dashed curve marked plus: [17] sub-sampling $\asymp n$ nodes; magenta dashed line with square markers: ASE plug-in bootstrap in [93].

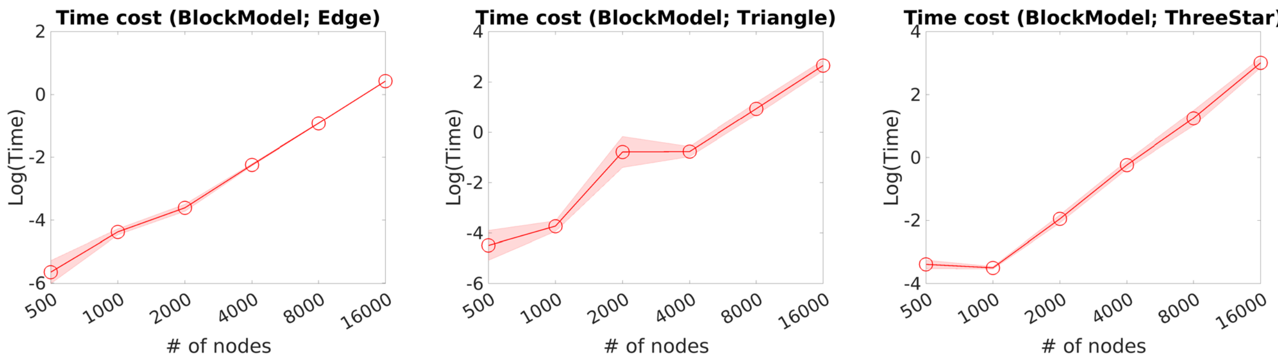

Fig 5: Scalability of our method on large networks. Bootstrap methods [17, 61, 93] with $n_{\text{boot}} = 200$ bootstrap iterations did not finish in 24 hours, thus are not shown. In all experiments, our method took less than 20 seconds at the longest to run.

Figure 5 shows the results. Notably, our method shows a clear uniform slope in their log-time cost growth rates for all the three motifs with $r = 2, 3, 4$, respectively. This echoes our discussion in Section 6 that some "nicely shaped" motifs only cost $O(n^2)$ or $O(n^3)$ to count, regardless of motif size r. On the other hand, we recognize that counting a large and "irregularly-shaped" motif could cost significantly more time.

6. Discussion. Our results do not cover the case where $g_1(X_1)$ is lattice and $\rho_n \asymp 1$. An ad-hoc remedy is to simply introduce artificial missing links by sparsifying A:

$$\widetilde{A}_{ij} := \widetilde{A}_{ji} := \begin{cases} A_{ij} = A_{ji}, & \text{with probability } 1 - \widetilde{\rho}_n \\ 0, & \text{with probability } \widetilde{\rho}_n \end{cases}$$

where we set $\widetilde{\rho}_n \asymp (\log n)^{-1}$. One can then make inferences about the population network moment $\widetilde{\rho}_n \cdot \mu_n$ using $\widetilde{A}$ as the input data (notice $\widetilde{\rho}_n$ is known). This reinstates the $(\log n)^{-1}$ sparsification that we need to overcome the latticeness at the price of a very minor information loss.

The Edgeworth expansion we derived for Bernoulli $A_{ij}|W_{ij}$ distributions can be readily extended to general weighted networks formulated by

$$A_{ij} := W_{ij} + \varepsilon_{ij},$$

where ε_{ij} may either depend on W_{ij} or not. A distinct feature of our setting is that the edge-wise observational errors are a contributing component of $\widehat{T}_n$ that smooths the distribution. In contrast to matrix estimation problems, where such noise is to be suppressed [33, 133], a moderate amount of tailedness can strengthen the smoothing effect in $A|W$ and might improve finite sample performances. Notice that similar to [17, 61, 93], throughout this paper, we work under the assumption inherited from well-known network analysis literature including [19, 34, 56, 40] that $\rho_n \cdot f(\cdot) \in [0,1]$, which also yields the boundedness of $h(\cdot)$. Thus, the bounded-moment conditions in the classical literature of Edgeworth expansions for noiseless U-statistics would be satisfied. There are at least two directions of potential relaxations: relaxing the boundedness of the distribution of $A_{ij}|W_{ij}$ and study a weighted network, or consider unbounded graphons like that in [26]. The extension of our algorithm and analysis to some light-tailed $A_{ij}|W_{ij}$ distributions is straightforward. For instance, our proofs remain valid for weighted network models with bounded graphon and an sub-exponential edge error distribution, where $W_{ij} \asymp \mathrm{Var}(\varepsilon_{ij}|W_{ij}) \asymp \rho_n$, by simply replacing Bernstein's inequality by generalized Hoeffding's inequality (Theorem 1.2.2 in [126]).

On the other hand, in fact there is a simple universal strategy to handle heavy-tailed A_{ij} distributions, regardless of whether this is due to a heavy-tailed $A_{ij}|W_{ij}$ distribution, or an unbounded graphon such as $f(x,y) = (xy)^{-\alpha}$ for $\alpha \in (0,1)$ in [26], or even both. As pointed out in [111], we can perform a one-to-one transformation, such as the widely-used *sigmoid* or *tanh* functions in machine learning, on each A_{ij}, tame it into a bounded $\mathcal{T}(A_{ij})$, and work with the transformed data. This also guarantees that the population network moments of the transformed network are always well-defined.

This paper focuses on studying the marginal randomness in A jointly contributed by the randomness in W and $A|W$. In this study, we take the sparsity-scaled graphon $\rho_n \cdot f$ as the population and the graphon feature μ_n as the ultimate inference goal. Our approach is nonparametric and directly approximates $F_{\widehat{T}_n}$ *without requiring* a graphon estimation $\widehat{W}$. If one regards W as the population and wants to make inference for U_n, she would need a CDF approximation to $(\widehat{U}_n - U_n)|X_1, \ldots, X_n$. This distribution is asymptotically normal as has been described by (3.12) in our Lemma 3.1-(b). However, estimating the normal variance typically *requires* a graphon estimation $\widehat{W}$[30]. Meanwhile, a practically meaningful graphon estimation would typically require that f is smooth and/or low-rank, see [136, 134]. In other words, the bootstrapping of $\widehat{T}_n|X_1, \ldots, X_n$ would (seemingly unavoidably) be a *parametric* bootstrap. In view of Lemma 3.1-(b), asympotically

$$(\rho_n \cdot n)^{1/2} \cdot \frac{\widehat{U}_n - U_n}{\sigma_n} \overset{d}{\approx} N(0, \sigma_w^2) \tag{6.1}$$

given $X_1, \ldots, X_n$, where recall that $\sigma_w \asymp 1$. However, the minimax rate for sparse graphon estimation (see [58, 85]) is

$$\text{Rescaled MSE: } (\rho_n \cdot n)^{-2} \cdot \|\widehat{W} - W\|_F^2 \asymp (\rho_n \cdot n)^{-1} \cdot \log n$$

If we use this error bound to control the estimation error of σ_w^2, then this yields an error of $|\widehat{\sigma}_w^2 - \sigma_w^2| \asymp n^{-1}\|\widehat{W} - W\|_F \asymp \rho_n^{1/2} \cdot n^{-1/2} \cdot \log^{1/2} n$. This error may dominate the $n^{-1/2}$ correction term in an Edgeworth expansion even for dense networks (e.g., Cramer's condition holds and $\rho_n \asymp 1$). Moreover, the minimax graphon estimation rate has not yet been achieved by any polynomial-time algorithm (see [136, 134] for comments), and using a practically feasible $\widehat{W}$ would cause an error $\gg n^{-1/2}$, ignoring ρ_n and log. Therefore, it might be challenging to accurately approximate the distribution of the LHS of (6.1) beyond asymptotic normality. Our observation here echos the common practice in network bootstrap literature [17, 61, 93] that they unanimously focus on the marginal distribution of $\widehat{U}_n$, rather than $(\widehat{U}_n - U_n)|X_1, \ldots, X_n$[31].

A retrospection on our simulation setting provides an interesting insight. In fact, the population Edgeworth expansion provides a much more efficient Monte Carlo procedure for simulating the true distribution $F_{\widehat{T}_n}$. Indeed, estimating ξ_1, $\mathbb{E}[g_1^3(X_1)]$ and $\mathbb{E}[g_1(X_1)g_1(X_2)g_2(X_1, X_2)]$ with $n_{\mathrm{MC}} \asymp n$ Monte Carlo samples yields a CDF approximation rate of $O\left(\mathcal{M}(\rho_n, n; R)\right) = o(n^{-1/2})$ when ρ_n satisfies the conditions of Theorem 3.1. This is much more efficient than the empirical CDF, which requires $n_{\mathrm{MC}} \succeq n^2$ to achieve the same accuracy order.

In the application of our results, we focus on node sampling network bootstraps. It is an interesting future work to investigate the higher-order accuracy properties of other schemes,

[30]The expression of σ_w^2 contains W_{ij}^2 terms originated from "$W_{ij}(1 - W_{ij})$" terms, which could not be estimated without a graphon estimation.

[31]For example, in Levin and Levina [93], the authors used a low-rank decomposition of A, which directly leads to an estimated $\widehat{W}$. But they also solely focused the marginal distribution of U_n (in our notation system).

such as sub-graph sampling [17] and (artificially) weighted bootstrap [93]. Also comprehensive numerical comparisons of different schemes under various settings would certainly be interesting for practitioners. As a closely related point, this paper studies the *complete* noisy U-statistic, "complete" in the sense that $(i_1, \ldots, i_r)$ ranges over all $\binom{n}{r}$ possible r-tuples. As one of the anonymous referees pointed out, evaluating the moment corresponding to an r-node motif would cost $O(n^r)$, which is still expensive for large n. Even for sparse networks, the counting may still need $O(\rho_n^{r-1} \cdot n^r)$ time using cutting-edge algorithms, see Section III.A of [2]. To accelerate the computation, papers [22, 105, 36, 87, 121] investigated this topic for the conventional noiseless U-statistic setting and formulated the Edgeworth expansion for "incomplete" U-statistics. They study noiseless incomplete U-statistics, and [17] proposed a "subgraph subsampling" scheme (their scheme (a)) that computes noisy incomplete U-statistics, which we call $\widehat{U}_n^{(\text{Incomplete})}$ for the network setting. Formally, define

$$\widehat{U}_n^{(\text{Incomplete})} := \frac{\sum_{1 \leqslant i_1 < \cdots < i_r \leqslant n} I_{i_1,\ldots,i_r} \cdot h(A_{i_1,\ldots,i_r})}{\sum_{1 \leqslant i_1 < \cdots < i_r \leqslant n} I_{i_1,\ldots,i_r}}$$

where $I_{i_1,\ldots,i_r}$'s are random variables indepedent of the network data. These $I_{i_1,\ldots,i_r}$'s can be i.i.d. Bernoulli, multinomial (if a given proportion of sub-sampled motifs is desired), or other reasonable sampling scheme distributions. It would be an interesting future research to carefully explore and quantify the self-smoothing effect for $\widehat{U}_n^{(\text{Incomplete})}$.

On the other hand, however, some particular motifs, such as *cycles*, *stars* and *wheels*, can be very efficiently evaluated, and the computational complexity may remain at most $O(n^3)$, instead of $O(n^r)$. For instance, the $\widehat{U}_n$ for *star* motifs can be approximately counted with ignorable error in just $O(n^2)$ time by averaging over $\{A^r_{(i,:)}\}_{i=1,\ldots,n}$. Another example is that a (k, ℓ)-*wheel* (see [20] for definition) can be evaluated in at most $O(n^3)$ time using the sample version of $Q(R)$ in Equation (2.9) in [20]. More readings along this line include [98], which provides detailed formula table for parallel computing up to $r = 4$ motifs, and [4] that studies fast-counting triangles in very large graphs. A recent paper [38] points out another promising direction of distributed computation.

Code. The MATLAB code for our method (empirical Edgeworth expansion) is available at `https://github.com/yzhanghf/NetworkEdgeworthExpansion`.

Acknowledgments. We deeply appreciate the insightful and constructive comments from the Editor, the Associate Editor and the anonymous referees that lead to significantly enriched scientific contents and much improved presentation of this paper. We thank Professor Bing-Yi Jing, Professor Junhui Wang and Professor Ji Zhu for helpful comments; Vincent Q. Vu for helpful feedback; Linjun Zhang and Yunpeng Zhao for constructive suggestions; and Keith Levin and Tianxi Li for discussions on network bootstraps. Dong Xia's research was partially supported by Hong Kong RGC Grant ECS 26302019, GRF 16303320 and Adobe Research Gift Award.

SUPPLEMENTARY MATERIAL

Supplement for: "Edgeworth expansions for network moments" (URL to be added). The supplementary material contains: (1). Definition of σ_w in Lemma 3.1-(b); (2). All proofs; and (3). Additional simulation results and accompanying interpretations.

SUPPLEMENTAL MATERIAL FOR: EDGEWORTH EXPANSIONS FOR NETWORK MOMENTS

By Yuan Zhang[1] and Dong Xia[2]

[1]Department of Statistics, The Ohio State University, yzhanghf@stat.osu.edu

[2]Department of Mathematics, Hong Kong University of Science and Technology, madxia@ust.hk

7. Definition of σ_w in Lemma 3.1-(b). The formal definition of σ_w we present here would complete the statement of this lemma. To start, we express $h(A_{i_1,\ldots,i_r}) := \mathbb{1}_{[A_{i_1,\ldots,i_r} \sqsupseteq R]}$ more explicitly as a sum of indicator product-terms, in which, each term checks if $A_{i_1,\cdots,i_r} \geqslant R_\pi$, where the $\geqslant$ is entry-wise and R_π is defined as $(R_\pi)_{ij} := R_{\pi(i)\pi(j)}$ with π ranging over all permutations. To formalize this, let $\mathrm{Perm}(R) := \{\pi^{(\ell)}, \ell = 1, \ldots, L\}$ denote the permutation group of R, where $\pi^{(1)} = \mathrm{id}$ is the identity map and $\pi^{(\ell_1)}(R) \neq \pi^{(\ell_2)}(R)$ for any $\ell_1 \neq \ell_2$. For simplicity, for all $1 \leqslant k_1 < k_2 \leqslant r$, define

$$J^{(k_1,k_2)}(x) = \begin{cases} x & \text{if } R_{k_1 k_2} = 1 \\ 1 & \text{if } R_{k_1 k_2} = 0 \end{cases}$$

Then $h(A_{i_1,\ldots,i_r})$ can be formally represented as

$$h(A_{i_1,\ldots,i_r}) = \sum_{\ell=1}^{L} \mathbb{1}_{\left[A_{i_1,\ldots,i_r} \geqslant R_{\pi^{(\ell)}}\right]} = \sum_{\ell=1}^{L} \prod_{1 \leqslant k_1 < k_2 \leqslant r} J^{\left(\pi^{(\ell)}(k_1), \pi^{(\ell)}(k_2)\right)}\left(A_{i_{k_1}, i_{k_2}}\right)$$

Define

$$\mathfrak{E}^{(\ell)}_{\{i_1,\ldots,i_r\},j_1,j_2} := J^{\left(\pi^{(\ell)}(j_1),\pi^{(\ell)}(j_2)\right)}\left(W_{i_{j_1}, i_{j_2}}\right) \tag{7.1}$$

$$\mathfrak{S}^{(\ell)}_{j_1,j_2} := \mathrm{Sign}\left\{J^{\left(\pi^{(\ell)}(j_1),\pi^{(\ell)}(j_2)\right)}\left(W_{i_{j_1}, i_{j_2}}\right)\right\} \tag{7.2}$$

where

$$\mathrm{Sign}(J) := \frac{\mathrm{d}J(x)}{\mathrm{d}x} = \begin{cases} +1 & \text{if } J(x) = x \\ 0 & \text{if } J(x) = 1 \end{cases}$$

and define

$$\widehat{\Theta}_{ij} := \frac{2r(r-1)}{\sigma_n \cdot \binom{n-2}{r-2}} \sum_{\substack{1 \leqslant i_1 < \cdots < i_r \leqslant n \\ \{i,j\} \subseteq \{i_1,\ldots,i_r\}}} \sum_{\ell=1}^{L} \left\{ \prod_{\substack{1 \leqslant j_1 < j_2 \leqslant r \\ (i_{j_1}, i_{j_2}) \neq (i,j)}} \mathfrak{E}^{(\ell)}_{\{i_1,\ldots,i_r\},j_1,j_2} \right\} \cdot \mathfrak{S}^{(\ell)}_{\substack{j_1', j_2': \\ (i_{j_1'}, i_{j_2'}) = (i,j)}}$$

Define σ_w as follows

$$\sigma_w^2 := \frac{\rho_n \cdot n}{\binom{n}{2}^2} \sum_{1 \leqslant i < j \leqslant n} \widehat{\Theta}_{ij}^2 \cdot W_{ij}(1 - W_{ij}) \tag{7.3}$$

This completes the statement of Lemma 3.1-(b).

8. Proofs.

8.1. *Bernstein-type concentration bound for multilinear polynomials of centered error terms.* Our proof would need the main result Theorem 1.3 of [116], which is also used in the proofs in [84]. To state this theorem, we first define a few preliminary quantities.

DEFINITION 8.1 (See Section 1.1 of [116]). *A* hypergraph *is formed by a node set $\mathcal{V}(H) := \{1, \ldots, N\} = [N]$ and a set $\mathcal{H}(H)$ of hyperedges, where a hyperedge $\mathfrak{h}$ of degree $\mathfrak{q}$ is defined to be a subset of nodes $\mathcal{V}(\mathfrak{h}) \subset \mathcal{V}(H)$ satisfying $|\mathcal{V}(\mathfrak{h})| \leqslant \mathfrak{q}$. We study the following multilinear polynomial*

$$\mathfrak{f}(\mathfrak{X}) := \sum_{\mathfrak{h}\in\mathcal{H}(H)} \mathfrak{W}_{\mathfrak{h}} \prod_{\mathfrak{v}\in\mathcal{V}(\mathfrak{h})} \mathfrak{X}_{\mathfrak{v}}$$

where $\mathfrak{X} = (\mathfrak{X}_1, \cdots, \mathfrak{X}_N)$ and $\mathfrak{W}_{\mathfrak{h}}$ is an weight multiplier on each hyperedge $\mathfrak{h}$. Suppose on each node we have a random variable, $Y := \{Y_1, \ldots, Y_N\}$ and a natural number $\mathfrak{r} \geqslant 0$. Let $\mathfrak{W}$ denote the set of all edge weights. We define:

$$\Xi_{\mathfrak{r}} := \Xi_{\mathfrak{r}}(Y, H, \mathfrak{W}) := \max_{\mathcal{S}\subseteq[N]:|\mathcal{S}|=\mathfrak{r}} \left(\sum_{\mathfrak{h}\in\mathcal{H}(H):\mathcal{S}\subseteq\mathcal{V}(\mathfrak{h})} |\mathfrak{W}_{\mathfrak{h}}| \prod_{\mathfrak{v}\in\mathcal{V}(\mathfrak{h})\setminus\mathcal{S}} \mathbb{E}[|Y_{\mathfrak{v}}|] \right) \tag{8.1}$$

where to avoid symbol conflict we replaced "μ" in [116] by "Ξ".

Next we cite the main assumption.

DEFINITION 8.2. *A random variable Z is called* central moment bounded *with parameter $\mathcal{L} > 0$, if for any integer $i \geqslant 1$, we have*

$$\mathbb{E}\left[|Z - \mathbb{E}[Z]|^i\right] \leqslant i\mathcal{L} \cdot \mathbb{E}\left[|Z - \mathbb{E}[Z]|^{i-1}\right]$$

Now we are ready to cite their main result.

THEOREM 8.1 (Theorem 1.3 of [116], also cited as Lemma 15 in [84]). *Suppose all $Y_1, \ldots, Y_N$ are independent (but not necessarily identically distributed) and they all satisfy the central moment bounded condition with a common parameter $\mathcal{L}$. Then we have*

$$\mathbb{P}\big(|\mathfrak{f}(Y) - \mathbb{E}[\mathfrak{f}(Y)]| \geqslant u\big)$$
$$\leqslant e^2 \cdot \max\left\{ \exp\left\{ -\frac{u^2}{C^{\mathfrak{q}} \cdot \mathrm{Var}(\mathfrak{f}(Y))} \right\}, \max_{1\leqslant\mathfrak{r}\leqslant\mathfrak{q}} \exp\left\{ -\left(\frac{u}{\Xi_{\mathfrak{r}} \mathcal{L}^{\mathfrak{r}} C^{\mathfrak{q}}} \right)^{1/\mathfrak{r}} \right\} \right\} \tag{8.2}$$

where C is a universal constant.

8.2. *Proof of Lemma 3.1.*

8.2.1. *Proof of Lemma 3.1-(a).* By the decomposition in [96], we have

$$\sigma_n^2 = \frac{r^2 \xi_1^2}{n} + O(\rho_n^{2s} \cdot n^{-2})$$

Therefore, $\sigma_n \asymp n^{-1/2} \cdot \xi_1 \asymp \rho_n^s \cdot n^{-1/2}$. Combining this fact with the Hoeffding's decomposition of $U_n - \mu_n$ in (3.1), we have

$$\frac{U_n - \mu_n}{\sigma_n} = \frac{\frac{r}{n}\sum_{i=1}^n g_1(X_i) + \frac{r(r-1)}{n(n-1)}\sum_{1\leqslant i<j\leqslant n} g_2(X_i, X_j) + \widetilde{R}_{3:r}}{\frac{r\xi_1}{\sqrt{n}} + O(\rho_n^s \cdot n^{-3/2})}$$

where

$$\widetilde{R}_{3:r} := \binom{n}{3}^{-1}\binom{n-3}{r-3}\sum_{1\leqslant i_1<i_2<i_3\leqslant n} g_3(X_{i_1}, X_{i_2}, X_{i_3})$$

$$+ \sum_{k=4}^r \binom{n}{k}^{-1}\binom{n-k}{r-k}\sum_{1\leqslant i_1<\cdots<i_k\leqslant n} g_k(X_{i_1}, \ldots, X_{i_k})$$

$$=: \widetilde{R}_3 + \widetilde{R}_{4:r} \tag{8.3}$$

and we also recall the definitions of $U_n^{\#}$ and Δ_n from (3.7) and the $O(\rho_n^s \cdot n^{-3/2})$ remainder control on the denominator is due to

$$\sigma_n = \frac{r\xi_1}{\sqrt{n}}\sqrt{1 + O(n^{-1})} = \frac{r\xi_1}{\sqrt{n}} + O(\rho_n^s \cdot n^{-3/2}).$$

Recall that for simplicity, throughout this paper we assume f is bounded, which implies the boundedness of the induced kernel function $h(\cdot)$. Therefore, the moment conditions of Lemma 1 of [96] are satisfied. By Lemma 1 of [96], we know that $\mathbb{E}[|\widetilde{R}_{4:r}|] = O_p(\rho_n^s \cdot n^{-2})$, thus by the remark below Lemma 2 in [96], this term is also $\widetilde{O}_p(\rho_n^s \cdot n^{-3/2})$. Now for $\widetilde{R}_3$, using Theorem 1 in [97], we know that $\widetilde{R}_3 = \widetilde{O}_p(\rho_n^s \cdot n^{-3/2} \cdot \log^{3/2} n)$. Therefore, we have

$$\frac{U_n - \mu_n}{\sigma_n} = U_n^{\#} + \Delta_n(1 + O(n^{-1})) + \widetilde{O}_p(n^{-1} \cdot \log^{3/2} n)$$

$$= U_n^{\#} + \Delta_n + \widetilde{O}_p(n^{-1} \cdot \log^{3/2} n)$$

This completes the proof of Lemma 3.1-(a).

Note that we use $\widetilde{R}$, $\breve{R}$ and $\mathring{R}$ to denote the remainder terms, where the "R" means "remainder". The properties of $\widetilde{R}$, $\breve{R}$ and $\mathring{R}$ certainly depend on the shape of the motif R, where we inherit the tradition of using "R" to represent the motif from past network moment method literature [20], but $\widetilde{R}$, $\breve{R}$ and $\mathring{R}$ are distinct notions from R.

8.2.2. *Proof of Lemma 3.1-(b).* We have

$$\binom{n}{r} \cdot \widehat{U}_n = \sum_{1\leqslant i_1<\cdots<i_r\leqslant n} h(A_{i_1,\ldots,i_r})$$

$$= \sum_{1\leqslant i_1<\cdots<i_r\leqslant n}\left\{\sum_{\ell=1}^L \prod_{1\leqslant j_1<j_2\leqslant r}\left(\mathfrak{E}^{(\ell)}_{\{i_1,\ldots,i_r\},j_1,j_2} + \mathfrak{S}^{(\ell)}_{j_1,j_2} \cdot \eta_{i_{j_1},i_{j_2}}\right)\right\}$$

$$=: \sum_{1\leqslant k_1<k_2\leqslant n} \widetilde{\Theta}_{k_1,k_2} \cdot \eta_{k_1,k_2} + \sum_{1\leqslant i_1<\cdots<i_r\leqslant n}\sum_{\ell=1}^L \prod_{1\leqslant j_1<j_2\leqslant r} \mathfrak{E}^{(\ell)}_{\{i_1,\ldots,i_r\},j_1,j_2} + \mathring{R}$$

$$= \sum_{1\leqslant k_1<k_2\leqslant n} \widetilde{\Theta}_{k_1,k_2} \cdot \eta_{k_1,k_2} + \binom{n}{r} \cdot U_n + \mathring{R}, \tag{8.4}$$

where we denote

$$\eta_{i,j} = A_{ij} - W_{ij},$$

$$\widetilde{\Theta}_{k_1,k_2} := \sum_{\substack{1\leqslant i_1<\cdots<i_r\leqslant n\\ \{k_1,k_2\}\subseteq\{i_1,\ldots,i_r\}}} \sum_{\ell=1}^{L} \left(\prod_{\substack{1\leqslant j_1<j_2\leqslant r\\ (i_{j_1},i_{j_2})\neq(k_1,k_2)}} \mathfrak{E}^{(\ell)}_{\{i_1,\ldots,i_r\},j_1,j_2} \right) \mathfrak{S}^{(\ell)}_{\substack{j_1',j_2':\\ (i_{j_1'},i_{j_2'})=(k_1,k_2)}},$$

where we recall the definitions of $\mathfrak{E}$ and $\mathfrak{S}$ from (7.1) and (7.2), respectively, and $\mathring{R} := \binom{n}{r}\sigma_n \cdot \breve{R}_n$ is the remainder that contains all unmentioned terms. Referring to the later formal definition of $\breve{\Delta}_n$ in (8.19) and recalling the definition of $\widehat{\Delta}_n$ in (3.7), one can also easily verify that by definition $\widehat{\Delta}_n = \breve{\Delta}_n + \breve{R}_n$. For clarity, we first verify that the coefficient in front of η_{k_1,k_2} is indeed $\widetilde{\Theta}_{k_1,k_2}$. For each $\{i_1,\ldots,i_r\} : 1 \leqslant i_1 < \cdots < i_r \leqslant n$ and each ℓ, the term

$$\prod_{1\leqslant j_1<j_2\leqslant r} \left(\mathfrak{E}^{(\ell)}_{\{i_1,\ldots,i_r\},j_1,j_2} + \mathfrak{S}^{(\ell)}_{j_1,j_2} \cdot \eta_{j_1,j_2} \right)$$

contributes to the coefficient of η_{k_1,k_2} if and only if $\{k_1,k_2\} \subseteq \{i_1,\ldots,i_r\}$. Now if (j_1',j_2') is the index pair from $\{1,\ldots,r\}$ such that $(i_{j_1'},i_{j_2'}) = (k_1,k_2)$, then itself contributes a multiplicative factor of $\mathfrak{S}^{(\ell)}_{j_1',j_2'}$ and every other pair $(i_{j_1},i_{j_2}) \neq (k_1,k_2)$ among $\{1,\ldots,r\}$ contributes a multiplicative factor of $\mathfrak{E}^{(\ell)}_{\{i_1,\ldots,i_r\},j_1,j_2}$, both into the term:

$$\left(\prod_{\substack{1\leqslant j_1<j_2\leqslant r\\ (i_{j_1},i_{j_2})\neq(k_1,k_2)}} \mathfrak{E}^{(\ell)}_{\{i_1,\ldots,i_r\},j_1,j_2} \right) \mathfrak{S}^{(\ell)}_{\substack{j_1',j_2':\\ (i_{j_1'},i_{j_2'})=(k_1,k_2)}}$$

as an additive term in the expression of $\widetilde{\Theta}_{k_1,k_2}$. This confirms that the coefficient of η_{k_1,k_2} is indeed $\widetilde{\Theta}_{k_1,k_2}$.

The main content of this proof is to show the finite sample convergence rate of the linear part to its asymptotic distribution, and to bound the remainder $\mathring{R}$.

Concentration inequality for the remainder term $\mathring{R}$

In this part of the proof, our focus is to bound the remainder term $\mathring{R}$. Without loss of generality, we inspect the coefficient in front of the term

$$\eta_{\left(k_1^{(1)},k_2^{(1)}\right)} \cdots \eta_{\left(k_1^{(v)},k_2^{(v)}\right)}$$

where $\left(k_1^{(1)},k_2^{(1)}\right),\ldots,\left(k_1^{(v)},k_2^{(v)}\right)$ are mutually different pairs from the set of node pairs formed by the first r indices $\left\{\left(\widetilde{k}_1,\widetilde{k}_2\right) : \widetilde{k}_1 < \widetilde{k}_2, \{\widetilde{k}_1,\widetilde{k}_2\} \subseteq \{i_1,\ldots,i_r\}\right\}$. This coefficient can be denoted and explicitly expanded as follows

$$\widetilde{\Theta}_{\mathcal{K}:=\left\{\left(k_1^{(1)},k_2^{(1)}\right),\ldots,\left(k_1^{(v)},k_2^{(v)}\right)\right\}}$$

$$:= \sum_{\substack{1\leqslant i_1<\cdots<i_r\leqslant n\\ \left(\cup_{j=1}^{v}\{k_1^{(j)},k_2^{(j)}\}\right)\subset\{i_1,\ldots,i_r\}}} \sum_{\ell=1}^{L} \left(\prod_{\substack{1\leqslant j_1<j_2\leqslant r\\ (i_{j_1},i_{j_2})\notin\mathcal{K}}} \mathfrak{E}^{(\ell)}_{\{i_1,\ldots,i_r\},j_1,j_2} \right) \left(\prod_{\substack{(j_1',j_2'):\\ (i_{j_1'},i_{j_2'})\in\mathcal{K}}} \mathfrak{S}^{(\ell)}_{j_1',j_2'} \right)$$

$$(8.5) \qquad =: \sum_{\ell=1}^{L} \widetilde{\Theta}_{\mathcal{K}}^{(\ell)}$$

Here we note a crucially important property of $\widetilde{\Theta}_{\mathcal{K}}^{(\ell)}$ that it is nonzero if and only if all of the node pairs in $\mathcal{K}$ are edges in the ℓ-th permuted version of the motif $\pi^{(\ell)}(R)$. This will be the key for us to effectively bound $\widetilde{\Theta}_{\mathcal{K}}$ and $\widehat{\Delta}^{(v,p)}$ in (8.7). We now upper bound $\widetilde{\Theta}_{\left(k_1^{(1)},k_2^{(1)}\right),\ldots,\left(k_1^{(v)},k_2^{(v)}\right)}$ for all $v \geqslant 2$, and this is an important step in upper bounding $\mathring{R}$. Define p

$$p := \left|\{k_1^{(1)}, k_2^{(1)}\} \cup \cdots \cup \{k_1^{(v)}, k_2^{(v)}\}\right|$$

to be the number of distinct indexes among $\left(k_1^{(1)}, k_2^{(1)}\right), \ldots, \left(k_1^{(v)}, k_2^{(v)}\right)$. Clearly, for $v \geqslant 2$, we have

$$3 \leqslant p \leqslant r, \quad \text{and} \quad \frac{p}{2} \leqslant v \leqslant \begin{cases} p-1, & \text{for acyclic } R, \\ p(p-1)/2, & \text{for cyclic } R, \end{cases}$$

It suffices to bound inside part of the right hand side of (8.5) for each fixed set of indices $\{i_1, \ldots, i_r\}$ and ℓ, because multiplying such upper bound by $\binom{n-p}{r-p}$ gives an upper bound on $\widetilde{\Theta}_{\mathcal{K}}$, ignoring constant factors including L. For each fixed ℓ and given $i_1, \ldots, i_r$ and $\mathcal{K}$, we see that the number of $\left(k_1^{(j)}, k_2^{(j)}\right)$ that correspond to edges under the permutation mapping $\pi^{(\ell)}$ must be v, otherwise at least one $\mathfrak{S}$ term is zero and the summand is zero. So we have

$$\left| \prod_{\substack{1 \leqslant j_1 < j_2 \leqslant r \\ (i_{j_1}, i_{j_2}) \notin \mathcal{K}}} \mathfrak{E}^{(\ell)}_{\{i_1,\ldots,i_r\},j_1,j_2} \right| \asymp \rho_n^{s-v}$$

and consequently,

$$(8.6) \qquad \left|\widetilde{\Theta}_{\left(k_1^{(1)},k_2^{(1)}\right),\ldots,\left(k_1^{(v)},k_2^{(v)}\right)}\right| \leq \rho_n^{s-v} \cdot \binom{n-p}{r-p} \asymp \rho_n^{s-v} n^{r-p}$$

We can express the remainder term $\mathring{R}$ in terms of $\widetilde{\Theta}$ and η terms. To facilitate detailed discussion and bounding, we group these terms. Define

$$(8.7) \qquad \widehat{\Delta}^{(v,p)} := \sum_{\substack{\mathcal{K} \subseteq \{(k_1,k_2): 1 \leqslant k_1 < k_2 \leqslant n\} \\ \text{Unique nodes}(\mathcal{K}) = p \\ |\mathcal{K}| = v}} \left(\widetilde{\Theta}_{\mathcal{K}} \prod_{(k_1,k_2) \in \mathcal{K}} \eta_{k_1,k_2} \right)$$

to be the collection of the terms in the remainder $\mathring{R}$ corresponding to the product over v different η-terms with exactly p unique participating nodes in these η-terms' indexes. Then

$$(8.8) \qquad \mathring{R} = \sum_{\substack{\text{All possible } (v,p) \\ v \geqslant 2, p \geqslant 3}} \widehat{\Delta}^{(v,p)}$$

Obviously, v, p and the total number of possible (v,p) pairs are all universally bounded, because the motif R is fixed. Therefore, in order to bound $\mathring{R}$, it suffices to bound $\widehat{\Delta}^{(v,p)}$ for every (v,p) pair. We need to bound not only the asymptotic magnitude of $\widehat{\Delta}^{(v,p)}$, but also its tail probability. Notice that $\widehat{\Delta}^{(v,p)}$ is mean zero both conditional on W and marginally.

In order to bound its tail probability, it suffice to show a proper concentration inequality for $\widehat{\Delta}^{(v,p)}$ conditional on W.

For this goal, we are going to apply Theorem 8.1, which derives a Bernstein inequality for polynomials of independent centered random variables. Notice that $\widehat{\Delta}^{(v,p)}$ can be rewritten in the form of (8.1), where the nodes of the hypergraph $\mathcal{V}(H) = \{(i,j) : 1 \leqslant i < j \leqslant n\}$ are defined to be the following set of *node pairs* (notice that they are *not* nodes but node pairs in our network). Define the set of hyperedges $\mathcal{H}(H)$ as follows

$$\begin{aligned}
\mathcal{H}(H) := \Big\{ \mathcal{K} := \Big\{ \Big(k_1^{(1)}, k_2^{(1)}\Big), \ldots, \Big(k_1^{(v)}, k_2^{(v)}\Big) \Big\} \subseteq \{(i,j) : 1 \leqslant i < j \leqslant n\} \\
\text{s.t.} \left| \cup_{v'=1}^{v} \{k_1^{(v')}, k_2^{(v')}\} \right| = p, \text{ and} \\
\text{there exists } 1 \leqslant i_1 < \cdots < i_r \leqslant n \text{ and } 1 \leqslant \ell \leqslant L, \text{ s.t.} \\
\mathcal{K} \in \{(i_{k_1'}, i_{k_2'}) : 1 \leqslant k_1' < k_2' \leqslant r\}, \\
\text{and } \Big(\pi^{(\ell)}(R)\Big)_{k_1', k_2'} = 1, \text{ for all } (k_1', k_2') : (i_{k_1'}, i_{k_2'}) \in \mathcal{K} \Big\}
\end{aligned}$$

In other words, using the notation in Theorem 8.1, $\mathcal{H}(H)$ is the collection of all size-v subsets of $\mathcal{V}(H)$ that span across p nodes and are subset to some ℓth permuted version of the motif $\pi^{(\ell)}(R)$, edge weights being $\mathfrak{W}_{\mathfrak{h}} = \widetilde{\Theta}_{\mathcal{K}}$, and each individual node-wise random variable is $\{Y_{\mathfrak{v}'}\} := \Big\{\eta_{k_1^{(\mathfrak{v}')}, k_2^{(\mathfrak{v}')}}\Big\}$. Clearly, centered Bernoulli random variables satisfy the "bounded central moment" assumption with parameter $\mathcal{L} = 1$. In our context, $\mathfrak{q} = v$. In order to apply Theorem 8.1, now we bound the key quantities $\Xi_1, \ldots, \Xi_v$. For each $q' : 1 \leqslant q' \leqslant v$, bounding $\Xi_{q'}$ consists of two sub-tasks:

(i). Bounding $\prod_{\mathfrak{v}' \in \mathcal{V}(\mathfrak{h}) \setminus \mathcal{S}} \mathbb{E}[|\eta_{v'}|]$,
(ii). Bounding $\sum_{\mathfrak{h} \in \mathcal{H}(H) : \mathcal{S} \subseteq V(\mathfrak{h})} |\mathfrak{W}_{\mathfrak{h}}| \cdot \prod_{\mathfrak{v}' \in \mathcal{V}(\mathfrak{h}) \setminus \mathcal{S}} \mathbb{E}[|\eta_{v'}|]$,

where in both bounds, $\mathcal{S} \subseteq \mathcal{V}(H) : |\mathcal{S}| = q'$. Bounding (i) is easy since it is just a product over $v - q'$ independent η terms, each of which has an absolute expectation of ρ_n. We have

$$\prod_{\mathfrak{v}' \in \mathcal{V}(\mathfrak{h}) \setminus \mathcal{S}} \mathbb{E}[|\eta_{v'}|] = \prod_{\mathfrak{v}' \in \mathcal{V}(\mathfrak{h}) \setminus \mathcal{S}} \mathbb{E}\Big[\mathbb{E}[|\eta_{v'}| \Big| W]\Big] \leq \rho_n^{v-q'} \tag{8.9}$$

Now we bound (ii). This requires more detailed calculations to count the number of $\widetilde{\Theta}$ terms involved in the summation. It turns out the bound would differ for acyclic and cyclic motifs, which we discuss as follows.

- When R is acyclic, in the summation $\sum_{\mathfrak{h} \in \mathcal{H}(H) : \mathcal{S} \subseteq V(\mathfrak{h})} |\mathfrak{W}_{\mathfrak{h}}|$, we are summing over at most $p - q' - 1$ free indices. To see this fact, recall that in order for an individual summand to be nonzero, its corresponding hyperedge $\mathfrak{h}$, or equivalently, the corresponding $\mathcal{K}$, must be a subset of some permuted version of the motif R. Therefore the requirement that it must contain $\mathcal{S} : |\mathcal{S}| = q'$ would pin down at least $q' + 1$ indices, leaving us at most $p - q' - 1$ free indices. Therefore, recalling that $|\mathfrak{W}_{\mathfrak{h}}| := |\widetilde{\Theta}_{\mathcal{K}}| \leq \rho_n^{s-v} n^{r-p}$, we obtain the following bound for (1)

$$\begin{aligned}
\sum_{\mathfrak{h} \in \mathcal{H}(H) : \mathcal{S} \subseteq V(\mathfrak{h})} |\mathfrak{W}_{\mathfrak{h}}| \cdot \prod_{\mathfrak{v}' \in \mathcal{V}(\mathfrak{h}) \setminus \mathcal{S}} \mathbb{E}[|\eta_{v'}|] &\leq \binom{n}{p - q' - 1} \cdot \rho_n^{s-v} n^{r-p} \cdot \rho_n^{v-q'} \\
&\leqslant \rho_n^{s-q'} \cdot n^{r-q'-1} \qquad (8.10)
\end{aligned}$$

Since (8.10) holds for any $\mathcal{S} : |\mathcal{S}| = q'$, by the definition of $\Xi_{q'}$, we have

$$\frac{\Xi_{q'}}{\binom{n}{r} \cdot \sigma_n} \preceq \frac{\rho_n^{s-q'} \cdot n^{r-q'-1}}{\rho_n^s \cdot n^{r-1/2}} = (\rho_n \cdot n)^{-q'} \cdot n^{-1/2} \tag{8.11}$$

Notice that under the weak sparsity assumption $\rho_n = \omega(n^{-1})$ for acyclic R, the RHS of (8.11) is decreasing in q'. The interpretation of the result (8.11) says that in fact, our choice of "u" in the second term inside "max" in Theorem 8.1 for $\mathfrak{r} : 1 \leqslant \mathfrak{r} \leqslant \mathfrak{q} = v$ is bottlenecked by the case $\mathfrak{r} = 1$.

- Then we discuss the more complicated case that R is cyclic. Now, we consider those $\mathcal{S} \subset \mathcal{V}(H)$ whose numbers of unique nodes are $q'' + 1$ for some $q'' \in \{2, \cdots, p-1\}$ ($q'' = 1$ cannot form a cyclic R). For such $\mathcal{S}$, we have

$$\sum_{\substack{\mathfrak{h} \in \mathcal{H}(H) : \mathcal{S} \subseteq V(\mathfrak{h}) \\ \text{Unique nodes}(\mathcal{S}) = q''+1}} |\mathfrak{W}_{\mathfrak{h}}| \leqslant \binom{n}{p - q'' - 1} \cdot \rho_n^{s-v} n^{r-p} \leqslant \rho_n^{s-v} \cdot n^{r-q''-1}$$

since we have $p - q'' - 1$ free indices to sum over. Meanwhile, regardless of the number of unique nodes in $\mathcal{S}$, we always have

$$\prod_{\mathfrak{v}' \in \mathcal{V}(\mathfrak{h}) \setminus \mathcal{S}} \mathbb{E}[|\eta_{\mathfrak{v}'}|] \preceq \rho_n^{v - |S|},$$

Now using the simple relationship $|S| \leqslant q''(q''+1)/2$, we have

$$\sum_{\mathfrak{h} \in \mathcal{H}(H) : \mathcal{S} \subseteq V(\mathfrak{h})} |\mathfrak{W}_{\mathfrak{h}}| \cdot \prod_{\mathfrak{v}' \in \mathcal{V}(\mathfrak{h}) \setminus \mathcal{S}} \mathbb{E}[|\eta_{\mathfrak{v}'}|]$$

$$= \sum_{\text{All possible } q''} \sum_{\substack{\mathfrak{h} \in \mathcal{H}(H) : \mathcal{S} \subseteq V(\mathfrak{h}) \\ \text{Unique nodes}(\mathcal{S}) = q''+1}} |\mathfrak{W}_{\mathfrak{h}}| \prod_{\mathfrak{v}' \in \mathcal{V}(\mathfrak{h}) \setminus \mathcal{S}} \mathbb{E}[|\eta_{\mathfrak{v}'}|]$$

$$\leqslant \sum_{q''} \binom{n}{p - q'' - 1} \cdot \rho_n^{s-v} n^{r-p} \cdot \rho_n^{v - q''(q''+1)/2}$$

$$\leqslant \sum_{q''} \rho_n^{s - q''(q''+1)/2} \cdot n^{r - q'' - 1} \tag{8.12}$$

Therefore, we have

$$\frac{\Xi_{q'}}{\binom{n}{r} \cdot \sigma_n} \preceq \max_{q'' : q''(q''+1)/2 \geqslant q'} \frac{\rho_n^{s - q''(q''+1)/2} \cdot n^{r - q'' - 1}}{\rho_n^s \cdot n^{r-1/2}}$$

$$= \max_{q'' : q''(q''+1)/2 \geqslant q'} \left(\rho_n^{-(q''+1)/2} \cdot n^{-1}\right)^{q''} \cdot n^{-1/2} \tag{8.13}$$

Recall that by definition $q'' \leqslant p - 1 \leqslant r - 1$. Under the weak sparsity assumption that $\rho_n = \omega(n^{-2/r})$, we know that $\rho_n^{-(q''+1)/2} \cdot n^{-1} \ll 1$, so the maximum asymptotic order on the RHS of (8.13) is achieved at the minimum possible q'' value of 2.

Now we have bounded the Ξ terms. In fact, as we will see, in Theorem 8.1, the concentration error bound terms due to Ξ's are dominated by the term due to variance. In order to apply Theorem 8.1, it only remains to bound $\operatorname{Var}\left(\widehat{\Delta}^{(v,p)}\right)$. We shall do this by bounding $\operatorname{Var}\left(\widehat{\Delta}^{(v,p)} | W\right)$ for each individual (v,p), since the number of such terms is a fixed number.

We have

$$\operatorname{Var}\left(\widehat{\Delta}^{(v,p)}|W\right) = \sum_{1\leqslant i_1<\ldots<i_p\leqslant n} \widetilde{\Theta}_{\mathcal{K}}^2 \cdot \operatorname{Var}\left(\prod_{(k_1,k_2)\in\mathcal{K}} \eta_{k_1,k_2}|W\right)$$

$$\preceq n^p \cdot \rho_n^{2s-2v} \cdot n^{2r-2p} \cdot \rho_n^v = \rho_n^{2s-v} \cdot n^{2r-p} \leqslant \rho_n^{2s-v} \cdot n^{2r-p} \tag{8.14}$$

where we used (8.6). Since $v \leqslant s$ and $p \geqslant 3$, this yields the following upper bound.

$$\frac{\left\{\operatorname{Var}\left(\widehat{\Delta}^{(v,p)}|W\right)\right\}^{1/2}}{\binom{n}{r}\cdot\sigma_n} \asymp \rho_n^{-s} \cdot n^{1/2-r} \cdot \left\{\operatorname{Var}\left(\widehat{\Delta}^{(v,p)}|W\right)\right\}^{1/2}$$

$$\preceq \left(\rho_n^{-s} \cdot n^{1/2-r}\right) \cdot \left(\rho_n^{s-v/2} \cdot n^{r-p/2}\right)$$

$$= \rho_n^{-v/2} \cdot n^{-(p-1)/2} \tag{8.15}$$

Next we discuss different upper bounds of the RHS of (8.15) based on different motif R shapes.

- **Case 1:** if R is acyclic, we have $v \leqslant p-1$. Combining this with the fact that $p \geqslant 3$ and Assumption (ii) of Lemma 3.1 that $\rho_n = \omega(n^{-1})$, we have

$$\rho_n^{-v/2} \cdot n^{-(p-1)/2} \leqslant (\rho_n \cdot n)^{-(p-1)/2} \leqslant (\rho_n \cdot n)^{-1} \tag{8.16}$$

- **Case 2:** if R is cyclic, we have $v \leqslant p(p-1)/2$. Combining this with the fact that $3 \leqslant p \leqslant r$ and Assumption (ii) of Lemma 3.1 that $\rho_n = \omega(n^{-2/r})$, we have

$$\rho_n^{-v/2} \cdot n^{-(p-1)/2} \leqslant (\rho_n^{-p(p-1)/2} \cdot n^{-(p-1)})^{1/2}$$

$$= \left(\rho_n^{-p/2} \cdot n^{-1}\right)^{(p-1)/2} \leqslant \rho_n^{-r/2} \cdot n^{-1} \tag{8.17}$$

Repeating this argument for every (v, p) pair, and plug (8.11), (8.13), (8.16) and (8.17) back into Theorem 8.1, we have

(8.18)

$$\mathbb{P}\left(\breve{R}_n := \frac{\mathring{R}}{\binom{n}{r}\cdot\sigma_n} \geqslant C \cdot \mathcal{M}(\rho_n, n; R)\right)$$

$$\preceq \begin{cases} \max\left\{\exp\left(-\frac{((\rho_n \cdot n)^{-1}\log^{1/2} n)^2}{(\rho_n\cdot n)^{-2}}\right), \exp\left(-\frac{(\rho_n\cdot n)^{-1}\log^{1/2} n}{(\rho_n\cdot n)^{-1}\cdot n^{-1/2}}\right)\right\}, & \text{for acyclic } R; \\ \max\left\{\exp\left(-\frac{((\rho_n^{-r/2}\cdot n^{-1})\log^{1/2} n)^2}{(\rho_n^{-r/2}\cdot n^{-1})^2}\right), \exp\left(-\frac{(\rho_n^{-r/2}\cdot n^{-1})\log^{1/2} n}{\rho_n^{-3} n^{-5/2}}\right)\right\}, & \text{for cyclic } R; \end{cases}$$

$$= O(n^{-1})$$

for a large enough universal constant C.

Asymptotic normality of the linear part $\breve{\Delta}_n$ and Berry-Esseen bound

Now, we focus on $\breve{\Delta}_n$, the linear part of $(\widehat{U}_n - U_n)/\sigma_n$ and show the uniform rate of its normal approximation. Recalling the definitions of $\breve{\Delta}_n$, $\widetilde{\Theta}_{ij}$ and $\widehat{\Theta}_{ij}$, ignoring the remainder term, we have

$$\breve{\Delta}_n := \text{Linear part of } \left(\frac{\widehat{U}_n - U_n}{\sigma_n}\right) = \frac{1}{\binom{n}{r}\cdot\sigma_n} \sum_{1\leqslant i<j\leqslant n} \widetilde{\Theta}_{ij} \cdot \eta_{ij}$$

$$=: \frac{1}{\binom{n}{2}} \sum_{1\leqslant i<j\leqslant n} \widehat{\Theta}_{ij}\cdot \eta_{ij} \tag{8.19}$$

We are going to show the asymptotic normality of $\check{\Delta}_n$ and its concentration speed by applying the Berry-Esseen bound for independent but differently-distributed random variables [39] conditioning on W. In this derivation, the key terms are the asymptotic orders of the second and third central moments of each individual $\widehat{\Theta}_{ij}\eta_{ij}$ term. We first show that with respect to the randomness in W, we have $\widehat{\Theta}_{ij} \asymp \rho_n^{-1}\cdot n^{1/2}$. Then when we condition on W and apply the generalized Berry-Esseen bound with respect to the randomness of A given W, we can think of $\widehat{\Theta}_{ij}$ as its asymptotic order $\rho_n^{-1}\cdot n^{1/2}$. Recall that

$$\prod_{\substack{1\leqslant j_1<j_2\leqslant r\\ (i_{j_1},i_{j_2})\neq(i,j)}} \mathfrak{E}^{(\ell)}_{\{i_1,\dots,i_r\},j_1,j_2}\cdot \mathfrak{S}^{(\ell)}_{\substack{j_1',j_2':\\ (i_{j_1'},i_{j_2'})=(i,j)}}$$

$$\asymp \begin{cases} \rho_n^{s-1}, & \text{if } \mathfrak{S}^{(\ell)}_{\substack{j_1',j_2':\\ (i_{j_1'},i_{j_2'})=(i,j)}}=1 \text{ or equivalently } \left(\pi^{(\ell)}(R)\right)_{j_1',j_2'}=1\\ 0, & \text{if } \mathfrak{S}^{(\ell)}_{\substack{j_1',j_2':\\ (i_{j_1'},i_{j_2'})=(i,j)}}=0 \text{ or equivalently } \left(\pi^{(\ell)}(R)\right)_{j_1',j_2'}=0 \end{cases}$$

We have

$$\sigma_n\cdot\widehat{\Theta}_{ij}\asymp\rho_n^{s-1}$$

This is because

$$\frac{\sigma_n\cdot\widehat{\Theta}_{ij}}{2r(r-1)} = \frac{1}{\binom{n-2}{r-2}} \sum_{\substack{1\leqslant i_1<\cdots<i_r\leqslant n\\ \{i,j\}\subset\{1,\dots,i_r\}}} \sum_{\ell=1}^{L} \left\{ \prod_{\substack{1\leqslant j_1<j_2\leqslant r\\ (j_1,j_2)\neq(i,j)}} \mathfrak{E}^{(\ell)}_{\{i_1,\dots,i_r\},j_1,j_2} \right\}\cdot \mathfrak{S}^{(\ell)}_{i,j} \tag{8.20}$$

Since for each given $\{i_1,\dots,i_r\}$ that contains $\{i,j\}$, the summation over ℓ ranges among all $\pi^{(\ell)}$ that keep (i,j) an edge in $\pi^{(\ell)}(R)$, so the outcome of this summation over ℓ is symmetric in $\{i_1,\dots,i_r\}\backslash\{i,j\}$. Consequently, $\widehat{\Theta}_{ij}$ is also symmetric in $\{1,2,\dots,n\}\backslash\{i,j\}$. Applying Hoeffding's decomposition to each $\widehat{\Theta}_{ij}$ viewed as a U-statistic with index set $\{1,\dots,n\}\backslash\{i,j\}$ and using [97] to bound the remainder, we have

(8.21)
$$\frac{\sigma_n\cdot\widehat{\Theta}_{ij}}{2r(r-1)} = \frac{\mathbb{E}\left[\sigma_n\cdot\widehat{\Theta}_{ij}\middle|X_i,X_j\right]}{2r(r-1)} + \frac{r-2}{n-2}\sum_{\substack{1\leqslant k\leqslant n\\ k\neq i,j}} \breve{g}_{1;i,j}(X_k) + \widetilde{O}_p(\rho_n^{s-1}\cdot n^{-1}\cdot\log^{3/2} n)$$

where

$$\breve{g}_{1;i,j}(X_k) := \mathbb{E}\left[\prod_{\substack{1\leqslant j_1<j_2\leqslant r\\ (i_{j_1},i_{j_2})\neq(i,j)}} \mathfrak{E}^{(\ell)}_{\{i_1,\dots,i_r\},j_1,j_2}\middle|X_k,X_i,X_j\right] - \frac{\mathbb{E}\left[\sigma_n\cdot\widehat{\Theta}_{ij}\middle|X_i,X_j\right]}{2r(r-1)}$$

where the indexes $i_1,\dots,i_r$ satisfy $\{i,j,k\}\subseteq\{i_1,\dots,i_r\}\subseteq\{1,\dots,n\}$. Since the linear part of a Hoeffding's decomposition are averaging over $\asymp n$ i.i.d. terms with $\mathbb{E}[\breve{g}_{1;i,j}(X_k)|X_i,X_j]=0, |\breve{g}_{1;i,j}(X_k)|=O(\rho_n^{s-1})\ a.s.$ and $\mathrm{Var}(\breve{g}_{1;i,j}(X_k)|X_i,X_j)\preceq\rho_n^{s-1}$, by Bernstein's inequality combined with a union bound, we have

$$\mathbb{P}\left(\max_{1\leqslant i<j\leqslant n}\frac{\sigma_n\cdot\left|\widehat{\Theta}_{ij}-\mathbb{E}\left[\widehat{\Theta}_{ij}\right]\right|}{2r(r-1)}\geqslant\rho_n^{s-1}\cdot t\,\middle|X_i,X_j\right)\leqslant C_1\binom{n}{2}\cdot\left\{e^{-C_2nt^2}+e^{-C_3nt}\right\}$$

which yields that conditioning on X_i, X_j, we have

$$\max_{1\leqslant i<j\leqslant n} \frac{\sigma_n \cdot \left|\widehat{\Theta}_{ij} - \mathbb{E}\left[\widehat{\Theta}_{ij}\right]\right|}{2r(r-1)} = \widetilde{O}_p\left(\rho_n^{s-1} \cdot n^{-1/2} \cdot \log n\right) \tag{8.22}$$

Since

$$\rho_n^{-(s-1)} \cdot \mathbb{E}\left[\sigma_n \cdot \widehat{\Theta}_{ij}\right] \asymp C > 0$$

for a universal constant C, when discussing the concentration of $\check{\Delta}_n$, it suffices to prove the Berry-Esseen bound for the asymptotic normality of $\check{\Delta}_n$ with respect to the randomness in $A|W$, conditioning on a "nicely-behaved" W such that $C/2 < \rho_n^{-(s-1)}\sigma_n \cdot \widehat{\Theta}_{ij} \asymp \rho_n^{-(s-1)}\sigma_n \cdot \widehat{\Theta}_{ij} < 3C/2$ holds for all $1 \leqslant i < j \leqslant n$ simultaneously, because the probability that W behaves "badly" is exponentially small and ignorable. We write

$$\frac{(\rho_n \cdot n)^{1/2} \cdot \check{\Delta}_n}{\sigma_w} = \sum_{1\leqslant i<j\leqslant n} \frac{(\rho_n \cdot n)^{1/2} \cdot \widehat{\Theta}_{ij}}{\sigma_w \cdot \binom{n}{2}} \cdot \eta_{ij} \tag{8.23}$$

where we notice that each individual coefficient in front of η_{ij} is at the order of $\rho_n^{-1/2} \cdot n^{-1}$. Using Theorem 2.1 of [39]

$$\left\| F_{\left.\frac{(\rho_n \cdot n)^{1/2} \cdot \check{\Delta}_n}{\sigma_w}\right|W}(u) - F_{N(0,1)}(u) \right\|_\infty \leqslant C\left\{0 + \sum_{1\leqslant i<j\leqslant n} \left(\frac{(\rho_n \cdot n)^{1/2} \cdot \widehat{\Theta}_{ij}}{\sigma_w \cdot \binom{n}{2}}\right)^3 \mathbb{E}\left[|\eta_{ij}|^3\right]\middle| W\right\}$$

$$\leq n^2 \cdot \rho_n^{-3/2} \cdot n^{-3} \cdot \rho_n \asymp \rho_n^{-1/2} \cdot n^{-1} \tag{8.24}$$

where we used

$$\mathbb{E}\left[|\eta_{ij}|^3 \,|W\right] = W_{ij}(1 - W_{ij})^3 + (1 - W_{ij})W_{ij}^3 \leqslant 2W_{ij} \asymp \rho_n$$

Recall that the above result was obtained under "nicely-bahaved" W, but the probability of "bad" W is exponentially small. Therefore, we have

$$\left\| F_{\left.\frac{(\rho_n \cdot n)^{1/2} \cdot \check{\Delta}_n}{\sigma_w}\right|W}(u) - F_{N(0,1)}(u) \right\|_\infty = \widetilde{O}_p(\rho_n^{-1/2} \cdot n^{-1}) \tag{8.25}$$

Combining (8.25) and (8.18) with Lemma 8.2 finishes the proof of Lemma 3.1-(b).

8.2.3. *Proof of Lemma 3.1-(c).* Define the following shorthand that will be used in not only this proof but also others

$$\widehat{a}_i := \frac{1}{\binom{n-1}{r-1}} \sum_{\substack{1\leqslant i_1<\ldots<i_{r-1}\leqslant n \\ i_1,\ldots,i_{r-1}\neq i}} h(A_{i,i_1,\ldots,i_{r-1}}) \tag{8.26}$$

$$a_i := \frac{1}{\binom{n-1}{r-1}} \sum_{\substack{1\leqslant i_1<\ldots<i_{r-1}\leqslant n \\ i_1,\ldots,i_{r-1}\neq i}} h(W_{i,i_1,\ldots,i_{r-1}}) \tag{8.27}$$

$$= \frac{1}{\binom{n-1}{r-1}} \sum_{\substack{1\leqslant i_1<\ldots<i_{r-1}\leqslant n \\ i_1,\ldots,i_{r-1}\neq i}} h(X_i, X_{i_1}, \ldots, X_{i_{r-1}})$$

A simple but useful property is as follows:

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{a}_i=\widehat{U}_n \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}a_i=U_n \tag{8.28}$$

To see (8.28), notice that

$$\sum_{i=1}^{n}\widehat{a}_i\cdot\binom{n-1}{r-1}=r\sum_{1\leqslant i_1<\cdots<i_r\leqslant n}h(A_{i_1,\dots,i_r})=r\cdot\binom{n}{r}\widehat{U}_n$$

because each $h(A_{i_1,\dots,i_r})$ is counted r times by $\widehat{a}_{i_1},\dots,\widehat{a}_{i_r}$, respectively, on the LHS. The relationship between a_i and U_n is verified exactly similarly.

Next, we start to decompose $\widehat{\delta}_n$. By definition, we have

$$\widehat{\delta}_n=\frac{\widehat{S}_n^2-\widehat{\sigma}_n^2}{\sigma_n^2}=\frac{\dfrac{n\widehat{S}_n^2}{r^2}-\dfrac{n\widehat{\sigma}_n^2}{r^2}}{\dfrac{n\sigma_n^2}{r^2}}$$

in which,

$$\begin{aligned}\frac{n\widehat{S}_n^2}{r^2}&=\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{a}_i-\widehat{U}_n\right)^2=\frac{1}{n}\sum_{i=1}^{n}\left\{(\widehat{a}_i-U_n)+\left(U_n-\widehat{U}_n\right)\right\}^2\\&=\frac{1}{n}\sum_{i=1}^{n}(\widehat{a}_i-U_n)^2+\frac{2}{n}\sum_{i=1}^{n}(\widehat{a}_i-U_n)\left(U_n-\widehat{U}_n\right)+\left(U_n-\widehat{U}_n\right)^2\\&=\frac{1}{n}\sum_{i=1}^{n}(\widehat{a}_i-U_n)^2-\left(U_n-\widehat{U}_n\right)^2 \qquad (8.29)\end{aligned}$$

By the earlier proof steps, we know that

$$\left(U_n-\widehat{U}_n\right)^2=O_p(\rho_n^{2s-1}n^{-2}) \tag{8.30}$$

According to the remark under Lemma 2 in [96], this term is $\widetilde{O}_p(\rho_n^{2s-1}\cdot n^{-1})$ and thus ignorable. We focus on decomposing the first term on the RHS of (8.29). We have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(\widehat{a}_i-U_n)^2&=\frac{1}{n}\sum_{i=1}^{n}\{(\widehat{a}_i-a_i)+(a_i-U_n)\}^2\\&=\frac{1}{n}\sum_{i=1}^{n}(\widehat{a}_i-a_i)^2+\frac{2}{n}\sum_{i=1}^{n}(\widehat{a}_i-a_i)(a_i-U_n)+\frac{1}{n}\sum_{i=1}^{n}(a_i-U_n)^2 \qquad (8.31)\end{aligned}$$

Term 3 on the RHS of (8.31) is the constituting part of $\widehat{\sigma}_n^2$, so we only need to bound the first two terms. The key component is to study $\widehat{a}_i-a_i$. Similar to the proof of part (b), starting from re-expressing the definition of $\widehat{a}_i$ and a_i, we have

$$\begin{aligned}\widehat{a}_i-a_i&=\frac{1}{\binom{n-1}{r-1}}\sum_{\substack{1\leqslant i_1<\cdots<i_r\leqslant n\\ i\in\{i_1,\dots,i_r\}}}\{h(A_{i_1,\dots,i_r})-h(W_{i_1,\dots,i_r})\}\\&=\frac{1}{\binom{n-1}{r-1}}\sum_{\text{All possible }(v,p)}\widehat{\Delta}^{(i;v,p)} \qquad (8.32)\end{aligned}$$

where recall that we use the shorthand $\mathcal{K} := \{(k_1^{(1)}, k_2^{(1)}), \ldots, (k_1^{(v)}, k_2^{(v)})\}$, and define

(8.33)
$$\widehat{\Delta}^{(i;v,p)} := \sum_{\substack{\mathcal{K}\subseteq\{(k_1,k_2):1\leqslant k_1<k_2\leqslant n\} \\ \text{Unique nodes}(\mathcal{K})=p \\ |\mathcal{K}|=v}} \widetilde{\Theta}_{\mathcal{K}}^{(i)} \prod_{(k_1,k_2)\in\mathcal{K}} \eta_{k_1,k_2}$$

(8.34)
$$\widetilde{\Theta}_{\mathcal{K}}^{(i)} := \sum_{\substack{1\leqslant i_1<\cdots<i_r\leqslant n \\ i\in\{i_1,\ldots,i_r\} \\ \mathcal{K}\subseteq\{(i_{j_1},i_{j_2}):1\leqslant j_1<j_2\leqslant r\}}} \sum_{\ell=1}^{L} \left(\prod_{1\leqslant j_1<j_2\leqslant r} \mathfrak{E}^{(\ell)}_{\{i_1,\ldots,i_r\},j_1,j_2} \right) \left(\prod_{(j_1',j_2'):(i_{j_1'},i_{j_2'})\in\mathcal{K}} \mathfrak{S}^{(\ell)}_{j_1',j_2'} \right)$$

Here we stress a crucial point that although in these definitions we always have $i \in \{i_1, \ldots, i_r\}$, the node i, however, might or might not appear in $\mathcal{K}$. This is because $\mathcal{K}$ is a subset of $\{(i_{j_1}, i_{j_2}) : 1 \leqslant j_1 < j_2 \leqslant r\}$. Conceptually assisted by this understanding, by counting the number of indexes over which the first summation in the definition of $\widetilde{\Theta}_{\mathcal{K}}^{(i)}$ is running, we have

$$\left|\widetilde{\Theta}_{\mathcal{K}}^{(i)}\right| \preceq \begin{cases} \rho_n^{s-v} \cdot n^{r-p}, & \text{if } i \in \text{Unique nodes}(\mathcal{K}) \\ \rho_n^{s-v} \cdot n^{r-p-1}, & \text{if } i \notin \text{Unique nodes}(\mathcal{K}) \end{cases} \tag{8.35}$$

Next, we separate the linear $\widehat{\Delta}^{(i;v,p)}$ terms, "linear" in the sense the are linear in $\eta_{(k_1,k_2)}$ terms, from those terms quadratic and higher degree in "η". The linear term corresponds to $(v, p) = (1, 2)$, and the higher degree terms correspond to $v \geqslant 2$ and $p \geqslant 3$. For the linear part, we have

$$\widehat{\Delta}^{(i;1,2)} = \sum_{1\leqslant j\leqslant n:j\neq i} \widetilde{\Theta}_{(i,j)}\eta_{i,j} + \sum_{\substack{1\leqslant j_1<j_2\leqslant n \\ j_1,j_2\neq i}} \widetilde{\Theta}_{(j_1,j_2)}\eta_{j_1,j_2} \tag{8.36}$$

Conditioned on W, applying Bernstein's inequality and (8.35) to the second term on the RHS of (8.36), respectively, we have

$$\widehat{\Delta}^{(i;1,2)} = \sum_{1\leqslant j\leqslant n:j\neq i} \widetilde{\Theta}_{(i,j)}\eta_{ij} + \widetilde{O}_p(\rho_n^{s-1/2} n^{r-2} \cdot \log n) \tag{8.37}$$

where the first term on the RHS of (8.37) is $\widetilde{O}_p(\rho_n^{s-1/2} n^{r-3/2} \cdot \log^{1/2} n)$.

Now we study the higher degree $\widehat{\Delta}^{(i;v,p)}$ terms. We are going to apply Theorem 8.1. We first upper bound "$\Xi_{q'}$" for all $q' = 1, \ldots, s$ as follows

- If R is acyclic:
  (i). If $i \in \mathcal{K}$: with "$|S| = q'$", we are summing over $(p-1) - q' - 1$ node indices in the summation $\sum_{\mathfrak{h}\in\mathcal{H}(H):\mathcal{V}(\mathfrak{h}):\mathcal{S}\subseteq\mathcal{V}(H)}$ – compared to the derivation of (8.11), here we have "$p-1$" instead of "p" because the index i is fixed and cannot vary in the summation. Therefore
  $$\sum_{\mathfrak{h}\in\mathcal{H}(H):\mathcal{V}(\mathfrak{h}):\mathcal{S}\subseteq\mathcal{V}(H)} |\mathfrak{W}_{\mathfrak{h}}| \preceq \rho_n^{s-v} n^{r-p} \cdot n^{p-q'-2} = \rho_n^{s-v} \cdot n^{r-q'-2}$$
  and consequently
  $$\Xi_{q'} \preceq \rho_n^{s-v} \cdot n^{r-q'-2} \cdot \rho_n^{v-q'} = \rho_n^{s-q'} \cdot n^{r-q'-2} \leqslant \rho_n^{s-1} \cdot n^{r-3}, \tag{8.38}$$
  under the weak sparsity assumption $\rho_n = \omega(n^{-1})$.

(ii). If $i \notin \mathcal{K}$: with "$|S| = q'$", we are summing over $p - q' - 1$ node indices in the summation $\sum_{\mathfrak{h}\in\mathcal{H}(H):\mathcal{V}(\mathfrak{h}):\mathcal{S}\subseteq\mathcal{V}(H)}$. But compared to the "$i \in \mathcal{K}$" case, here we lose an "n" factor in $|\widetilde{\Theta}_{\mathcal{K}}^{(i)}|$ according to (8.35). Therefore, we arrive at the identical upper bound for $\sum_{\mathfrak{h}\in\mathcal{H}(H):\mathcal{V}(\mathfrak{h}):\mathcal{S}\subseteq\mathcal{V}(H)}$ as the above case, namely, $\Xi_{q'} \preceq \rho_n^{s-1} \cdot n^{r-3}$ under the weak sparsity assumption $\rho_n = \omega(n^{-1})$.

- The proof for the case where R is cyclic can be obtained by revising the proof of (8.13). If $i \in \mathcal{K}$, we are summing over $(p-1) - q'' - 1$ instead of $p - q'' - 1$ node indices in $\sum_{\mathfrak{h}\in\mathcal{H}(H):\mathcal{V}(\mathfrak{h}):\mathcal{S}\subseteq\mathcal{V}(H)}$, if $i \notin \mathcal{K}$, then we sum over $p - q'' - 1$ node indices but will lose an n factor in the upper bound of $|\Theta_K^{(i)}|$ according to (8.35). Therefore, both cases would eventually lead to the same upper bound

$$\sum_{\mathfrak{h}\in\mathcal{H}(H):\mathcal{V}(\mathfrak{h}):\mathcal{S}\subseteq\mathcal{V}(H)} |\mathfrak{W}_{\mathfrak{h}}| \preceq \rho_n^{s-v} n^{r-p-1} \cdot n^{p-q''-1} = \rho_n^{s-v} \cdot n^{r-q''-2}$$

Similar to the proof of (8.13), it suffices to upper bound those $\Xi_{q'}$ where $q' = q''(q''+1)/2$, and we have

$$\Xi_{q''(q''+1)/2} \preceq \rho_n^{s-v} \cdot n^{r-q''-2} \cdot \rho_n^{v-q''(q''+1)/2} = \rho_n^{s-q''(q''+1)/2} \cdot n^{r-q''-2} \tag{8.39}$$

Same as before, the RHS is still monotone in q'' under the assumption $\rho_n = \omega(n^{-2/r})$ and thus it is bottlenecked by the $q'' = 1$ case.

Now in order to apply Theorem 8.1 to the higher degree $\widehat{\Delta}^{(i;v,p)}$ terms ($v \geqslant 2$ and $p \geqslant 3$), it only remains to calculate their conditional variances given W. Notice that given W, all $\widehat{\Delta}^{(i;v,p)}$ terms with different (v,p) configurations are mutually uncorrelated. We can bound each of them. Straightforward calculations show that

$$\begin{aligned}\mathrm{Var}\left\{\widehat{\Delta}^{(i;v,p)}|W\right\} &\preceq \underbrace{\binom{n-1}{p-1} \cdot \rho_n^{2s-2v} \cdot n^{2r-2p} \cdot \rho_n^{v}}_{\text{sum over } \mathcal{K}\text{-indexed terms: } i\in\mathcal{K}} + \underbrace{\binom{n}{p} \cdot \rho_n^{2s-2v} \cdot n^{2r-2p-2} \cdot \rho_n^{v}}_{\text{sum over } \mathcal{K}\text{-indexed terms: } i\notin\mathcal{K}}\\ &= O(\rho_n^{2s-v} \cdot n^{2r-p-1})\end{aligned}$$

- For acyclic R, if $\rho_n = \omega(n^{-1})$, we have

$$\begin{aligned}\mathrm{Var}\left\{\widehat{\Delta}^{(i;v,p)}|W\right\} &= O(\rho_n^{2s-v} \cdot n^{2r-p-1}) \leqslant O(\rho_n^{2s-(p-1)} \cdot n^{2r-p-1})\\ &\leqslant O(\rho_n^{2s} \cdot n^{2r-2} \cdot (\rho_n \cdot n)^{-2})\end{aligned} \tag{8.40}$$

- For cyclic R, if $\rho_n = \omega(n^{-2/r})$, we have

$$\begin{aligned}\mathrm{Var}\left\{\widehat{\Delta}^{(i;v,p)}|W\right\} &= O(\rho_n^{2s-v} \cdot n^{2r-p-1}) \leqslant O(\rho_n^{2s} \cdot n^{2r-1} \cdot \rho_n^{-p(p-1)/2} \cdot n^{-p})\\ &\leqslant O(\rho_n^{2s} \cdot n^{2r-2} \cdot (\rho_n^{-r/2} \cdot n^{-1})^2)\end{aligned} \tag{8.41}$$

Combining (8.38), (8.39), (8.40) and (8.41) with Theorem 8.1, we see that the sum of all higher degree $\widehat{\Delta}^{(i;v,p)}$ terms into $\widehat{\delta}_n$ is at the order of

$$\sum_{\substack{\text{All possible } (v,p):\\ v\geqslant 2,p\geqslant 3}} \widehat{\Delta}^{(i;v,p)} = \widetilde{O}_p(\rho_n^{s} \cdot n^{r-1} \cdot \mathcal{M}(\rho_n, n; R)) \tag{8.42}$$

Compared to the order of the linear $\widehat{\Delta}^{(i;v,p)}$ terms as the leading term on the RHS of (8.37), we see that the higher degree terms are ignorable.

Therefore, for the rest of the proof of Lemma 3.1-(c), we can replace $\widehat{a}_i - a_i$ by

$$\widehat{a}_i - a_i = \frac{1}{\binom{n-1}{r-1}} \left\{ \sum_{\substack{1\leqslant j\leqslant n\\ j\neq i}} \widetilde{\Theta}_{(i,j)}\eta_{ij} + \widetilde{O}_p(\rho_n^s \cdot n^{r-1} \cdot \mathcal{M}(\rho_n, n; R)) \right\}$$

$$(8.43)\qquad =: \frac{1}{n-1} \sum_{\substack{1\leqslant j\leqslant n\\ j\neq i}} \breve{\Theta}_{ij}\eta_{ij} + \widetilde{O}_p(\rho_n^s \cdot \mathcal{M}(\rho_n, n; R))$$

where

$$\breve{\Theta}_{ij} := \frac{r-1}{\binom{n-2}{r-2}} \widetilde{\Theta}_{(i,j)} \preceq \rho_n^{s-1}$$

according to (8.35).

Now we are ready to bound the first two terms on the RHS of (8.31) and finish the proof of Lemma 3.1-(c). For term 1, by Bernstein inequality and $\rho_n = \omega(n^{-1})$, we have

$$(8.44)\qquad \frac{1}{n}\sum_{i=1}^n (\widehat{a}_i - a_i)^2 = \widetilde{O}_p\big(\rho_n^{2s} \cdot \mathcal{M}(\rho_n, n; R)\big)$$

For term 2, recalling $|a_i - U_n| \preceq \rho_n^s$, we have

(8.45)
$$\frac{2}{n}\sum_{i=1}^n (a_i - U_n)(\widehat{a}_i - a_i) = \frac{2}{n(n-1)} \sum_{\substack{1\leqslant i\leqslant n\\ 1\leqslant j\leqslant n\\ i\neq j}} (a_i - U_n)\breve{\Theta}_{ij}\eta_{ij} + \widetilde{O}_p(\rho_n^{2s} \cdot \mathcal{M}(\rho_n, n; R))$$

conditioned on W. Applying Bernstein's inequality to the first on the RHS of (8.45) yields a bound of $\widetilde{O}_p(\rho_n^{2s-1/2} \cdot n^{-1}\log n)$. This completes the proof of Lemma 3.1-(c).

8.2.4. *Proof of Lemma 3.1-(d).* By definition, we have

$$\frac{n\widehat{\sigma}_n^2}{r^2} = \frac{1}{n}\sum_{i=1}^n \left\{ \frac{1}{\binom{n-1}{r-1}} \sum_{\substack{1\leqslant i_1<\cdots<i_{r-1}\leqslant n\\ i_1,\ldots,i_{r-1}\neq i}} h(W_{i,i_1,\ldots,i_{r-1}}) - U_n \right\}^2$$

$$= \frac{1}{n}\sum_{i=1}^n (a_i - U_n)^2 = \frac{1}{n}\sum_{i=1}^n \left\{(a_i - \mu_n)^2 + 2(a_i - \mu_n)(\mu_n - U_n) + (\mu_n - U_n)^2\right\}$$

$$= \frac{1}{n}\sum_{i=1}^n (a_i - \mu_n)^2 - (U_n - \mu_n)^2$$

Recalling Hoeffding's decomposition for U_n and applying Theorem 1 of [97] to bound the high-order canonical U-statistics, we have

$$(U_n - \mu_n)^2 = \left\{ \frac{r}{n}\sum_{i=1}^n g_1(X_i) + \widetilde{O}_p(\rho_n^s n^{-1} \cdot \log n) \right\}^2 = \widetilde{O}_p(\rho_n^{2s} n^{-1} \cdot \log n)$$

We focus on the first term. For notation convenience, define

$$\widetilde{a}_i := \mathbb{E}\left[h(X_i, X_{i_1}, \ldots, X_{i_{r-1}})|X_i\right] = g_1(X_i) + \mu_n$$

where $i_1,\dots,i_{r-1}\neq i$ are distinct indexes. We have

$$\frac{1}{n}\sum_{i=1}^{n}(a_i-\mu_n)^2=\frac{1}{n}\sum_{i=1}^{n}\left\{(a_i-\widetilde{a}_i)+(\widetilde{a}_i-\mu_n)\right\}$$

$$=\frac{1}{n}\sum_{i=1}^{n}(a_i-\widetilde{a}_i)^2+\frac{2}{n}\sum_{i=1}^{n}(a_i-\widetilde{a}_i)(\widetilde{a}_i-\mu_n)+\frac{1}{n}\sum_{i=1}^{n}(\widetilde{a}_i-\mu_n)^2 \tag{8.46}$$

First, we realize that term 3 on the RHS of (8.46) is simply

$$\frac{1}{n}\sum_{i=1}^{n}(\widetilde{a}_i-\mu_n)^2=\frac{1}{n}\sum_{i=1}^{n}g_1^2(X_i) \tag{8.47}$$

Now we focus on handling terms 1 and 2. The key part is to handle $a_i-\widetilde{a}_i$. By applying the Hoeffding's ANOVA decomposition of an arbitrary symmetric statistic (1.1)–(1.3) in [14] onto each single $h(X_i,X_{i_1},\dots,X_{i_{r-1}})$ term, we can see that

$$a_i-\widetilde{a}_i=\frac{1}{\binom{n-1}{r-1}}\sum_{\substack{1\leqslant i_1<\cdots<i_{r-1}\leqslant n\\ i_1,\dots,i_{r-1}\neq i}}\left\{h(X_i,X_{i_1},\dots,X_{i_{r-1}})-\mathbb{E}[h(X_i,X_{i_1},\dots,X_{i_{r-1}})|X_i]\right\}$$

$$=\frac{1}{\binom{n-1}{r-1}}\sum_{\substack{1\leqslant i_1<\cdots<i_{r-1}\leqslant n\\ i_1,\dots,i_{r-1}\neq i}}\left\{\sum_{k=1}^{r-1}\sum_{\substack{1\leqslant j_1<\cdots<j_k\leqslant n\\ \{j_1,\dots,j_k\}\subseteq\{i_1,\dots,i_{r-1}\}}}g_{k+1}(X_i,X_{j_1},\dots,X_{j_k})\right\}$$

$$=\frac{1}{\binom{n-1}{r-1}}\sum_{k=1}^{r-1}\binom{n-k-1}{r-k-1}\sum_{\substack{1\leqslant j_1<\cdots<j_k\leqslant n\\ j_1,\dots,j_k\neq i}}g_{k+1}(X_i,X_{j_1},\dots,X_{j_k}) \tag{8.48}$$

Now we apply Theorem 1 of [97] to the RHS of (8.48), we see that

$$a_i-\widetilde{a}_i=\frac{r-1}{n-1}\sum_{\substack{1\leqslant j\leqslant n\\ j\neq i}}g_2(X_i,X_j)+\widetilde{O}_p(\rho_n^s n^{-1}\cdot\log n) \tag{8.49}$$

$$=\widetilde{O}_p(\rho_n^s n^{-1/2}\cdot\log^{1/2}n) \tag{8.50}$$

Now we are ready to continue bounding the RHS of (8.46). Using (8.50), term 1 on the RHS of (8.46) is

$$(a_i-\widetilde{a}_i)^2=\widetilde{O}_p(\rho_n^{2s}n^{-1}\cdot\log n) \tag{8.51}$$

Using (8.49), term 2 on the RHS of (8.46) is

$$\frac{2}{n}\sum_{i=1}^{n}(a_i-\widetilde{a}_i)(\widetilde{a}_i-\mu_n)$$

$$=\frac{2}{n}\sum_{i=1}^{n}\left\{\frac{r-1}{n-1}\sum_{\substack{1\leqslant j\leqslant n\\ j\neq i}}g_2(X_i,X_j)+\widetilde{O}_p(\rho_n^s n^{-1}\cdot\log n)\right\}g_1(X_i)$$

$$=\frac{2(r-1)}{n(n-1)}\sum_{\substack{1\leqslant i\leqslant n\\ 1\leqslant j\leqslant n\\ i\neq j}}g_1(X_i)g_2(X_i,X_j)+\widetilde{O}_p(\rho_n^{2s}n^{-1}\cdot\log n) \tag{8.52}$$

Finally, combining (8.51), (8.52) and (8.47) completes the proof of Lemma 3.1-(d).

8.3. *Proof of Theorem 3.1.* We mainly prove for the case $\rho_n = O\left((\log n)^{-1}\right)$ without non-lattice condition. We will explain how this proof can be revised for the other case with Carmer's condition but without a ρ_n upper bound.

LEMMA 8.1 (Esseen's smoothing lemma ([52], Section XVI.3)). *For any distribution function F and a general function G that has universally bounded derivative and satisfy $G(-\infty) = 0, G(\infty) = 1$, we have*

$$\|F(u) - G(u)\|_\infty \leqslant C_1 \int_{-\gamma}^{\gamma} \left| \frac{\textit{Ch.f.}(F;t) - \textit{Ch.f.}(G;t)}{t} \right| \mathrm{d}t + \frac{C_2 \sup_u |G'(u)|}{\gamma} \tag{8.53}$$

for universal constants $C_1, C_2 > 0$, where Ch.f.$(G;t)$ is defined to be the characteristic function of G as follows

$$\textit{Ch.f.}(G;t) := \int_{-\infty}^{\infty} e^{\mathbb{i}tx} \mathrm{d}G(x)$$

Recall the definition of $\widetilde{T}_n$ from (3.9) that

$$\widetilde{T}_n = U_n^{\#} + \Delta_n - \frac{U_n^{\#}}{2}\delta_n \quad \text{and} \quad \widehat{T}_n = \widetilde{T}_n + \widehat{\Delta}_n + \widetilde{O}_p\big(\mathcal{M}(\rho_n, n; R)\big).$$

We define a random variable $\widetilde{\Delta}_n | W \sim N(0, (\rho_n \cdot n)^{-1}\sigma_\omega^2)$, that $\widetilde{\Delta}_n$ is conditionally independent of A, given W. By Lemma 3.1-(b), we have $\sup_{u \in \mathbb{R}} \big|F_{\widetilde{\Delta}_n}(u) - F_{\check{\Delta}_n}(u)\big| = O(\rho_n^{-1/2} \cdot n^{-1})$. We are going to show that

$$\left\| F_{\widehat{T}_n}(u) - F_{\widetilde{T}_n + \check{\Delta}_n}(u) \right\|_\infty = O\left(\mathcal{M}(\rho_n, n; R)\right) \tag{8.54}$$

$$\left\| F_{\widetilde{T}_n + \check{\Delta}_n}(u) - F_{\widetilde{T}_n + \widetilde{\Delta}_n}(u) \right\|_\infty = O(\rho_n^{-1/2} \cdot n^{-1}) \tag{8.55}$$

$$\left\| F_{\widetilde{T}_n + \widetilde{\Delta}_n}(u) - G_n(u) \right\|_\infty = O((\rho_n \cdot n)^{-1} + n^{-1} \log n) \tag{8.56}$$

where $G_n(u)$ is defined in (3.13). To proceed, we need the following smoothing lemma.

LEMMA 8.2. *Suppose we have random variables X, Y, Z satisfying*

$$X = Y + Z$$

such that the CDF of Y is smooth, and there exists a universal constant $0 < M < \infty$ such that $F_Y(u+a) - F_Y(u) \leqslant M \cdot a + O(\zeta_n)$ for any $u \in \mathbb{R}$ and $a > 0$. Also assume that $\mathbb{P}(|Z| \geqslant \widetilde{\zeta}_n) \leqslant n^{-1}$, that is, $Z = \widetilde{O}_p(\widetilde{\zeta}_n)$. We have

$$\|F_X(u) - F_Y(u)\|_\infty = O(\zeta_n + \widetilde{\zeta}_n + n^{-1})$$

Remark. We emphasize that Lemma 8.2 does *not* require any independence between X, Y and Z.

PROOF OF LEMMA 8.2. Since "$Y + Z > u$" implies the union of the following two events: "$Y > u - a$" and "$Z > a$", we have $1 - \mathbb{P}(Y + Z \leqslant u) \leqslant 1 - \mathbb{P}(Y \leqslant u - a) + \mathbb{P}(|Z| > a)$, which further implies that

$$\begin{aligned}
\mathbb{P}(Y + Z \leqslant u) &\geqslant \mathbb{P}(Y \leqslant u - a) - \mathbb{P}(|Z| > a) \\
&\geqslant \mathbb{P}(Y \leqslant u) - M \cdot a - O(\zeta_n) - \mathbb{P}(|Z| > a) \\
(\text{Setting } a = \widetilde{\zeta}_n) &\geqslant \mathbb{P}(Y \leqslant u) - O(\zeta_n + \widetilde{\zeta}_n + n^{-1})
\end{aligned}$$

On the other hand, we have

$$
\begin{aligned}
\mathbb{P}(Y+Z\leqslant u)=&\int_z \mathbb{P}(Y\leqslant u-z|Z=z)\mathrm{d}P_Z(z)\\
=&\int_{z:|z|\leqslant a}\mathbb{P}(Y\leqslant u-z|Z=z)\mathrm{d}P_Z(z)+\int_{z:|z|>a}\mathbb{P}(Y\leqslant u-z|Z=z)\mathrm{d}P_Z(z)\\
\leqslant&\int_z \mathbb{P}(Y\leqslant u+a|Z=z)\mathrm{d}P_Z(z)+\int_{z:|z|>a}1\,\mathrm{d}P_Z(z)\\
\leqslant&\mathbb{P}(Y\leqslant u+a)+\mathbb{P}(|Z|\geqslant a)
\end{aligned}
$$

Setting $a=\widetilde{\zeta}_n$, the RHS is upper bounded by $\mathbb{P}(Y\leqslant u)+O(\zeta_n+\widetilde{\zeta}_n+n^{-1})$. Combining the two inequalities proves Lemma 8.2. □

Now we return to the main proof of Theorem 3.1. Our proof would proceed as follows. We shall use Lemma 3.1-(b) to prove (8.55); then with the assistance of Lemma 8.2, we use (8.56) and (8.55) to prove (8.54); finally, we state the proof of (8.56) without needing (8.54) or (8.55).

- Proof of "Lemma 3.1-(b) $\Rightarrow$ (8.55)". Noticing that $\widetilde{T}_n$ does not depend on the random variations of $A|W$ given W, but it is determined if W is given, we have

$$
\begin{aligned}
F_{\widetilde{T}_n+\widecheck{\Delta}_n}(u)=&\mathbb{P}\left(\widetilde{T}_n+\widecheck{\Delta}_n\leqslant u\right)\\
=&\mathbb{E}\left[\mathbb{P}\left(\widetilde{T}_n+\widecheck{\Delta}_n\leqslant u|W\right)\right]\\
=&\mathbb{E}\left[\mathbb{P}\left(\widecheck{\Delta}_n\leqslant u-\widetilde{T}_n|W\right)\right]\\
\text{Lemma 3.1-(b) }=&\mathbb{E}\left[\mathbb{P}\left(\widetilde{\Delta}_n\leqslant u-\widetilde{T}_n|W\right)+\widetilde{O}_p(\rho_n^{-1/2}\cdot n^{-1})\right]\\
=&\mathbb{E}\left[\mathbb{P}\left(\widetilde{T}_n+\widetilde{\Delta}_n\leqslant u|W\right)\right]+O(\rho_n^{-1/2}\cdot n^{-1})\\
=&F_{\widetilde{T}_n+\widetilde{\Delta}_n}(u)+O(\rho_n^{-1/2}\cdot n^{-1})
\end{aligned}
$$

- Proof of "(8.55), (8.56) and Lemma 8.2 $\Rightarrow$ (8.54)". We set $Y=\widetilde{T}_n+\widecheck{\Delta}_n$ and $Z=\widehat{T}_n-Y$. We notice that by Lemma 3.1-(b), we have $Z=\widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)$ meaning that $\mathbb{P}\big(|Z|\geqslant C_1\mathcal{M}(\rho_n,n;R)\big)=O(n^{-1})$. Next we verify that Y satisfies the condition of Lemma 8.2, we notice that (8.56) implies that for any $u\in\mathbb{R}$ and $a>0$, we have

$$
\begin{aligned}
&F_{\widetilde{T}_n+\widecheck{\Delta}_n}(u+a)-F_{\widetilde{T}_n+\widecheck{\Delta}_n}(u)\\
\leqslant&\underbrace{\left|F_{\widetilde{T}_n+\widecheck{\Delta}_n}(u+a)-F_{\widetilde{T}_n+\widetilde{\Delta}_n}(u+a)\right|}_{\text{Bounded by (8.55)}}+\underbrace{\left|F_{\widetilde{T}_n+\widecheck{\Delta}_n}(u)-F_{\widetilde{T}_n+\widetilde{\Delta}_n}(u)\right|}_{\text{Bounded by (8.55)}}\\
&+\underbrace{\left|F_{\widetilde{T}_n+\widetilde{\Delta}_n}(u+a)-G_n(u+a)\right|}_{\text{Bounded by (8.56)}}+\underbrace{\left|G_n(u+a)-G_n(u)\right|}_{\sup_{u,n}|G_n'(u)|<\infty}\\
&+\underbrace{\left|F_{\widetilde{T}_n+\widetilde{\Delta}_n}(u)-G_n(u)\right|}_{\text{Bounded by (8.56)}}\\
\leqslant&C\cdot a+O(\rho_n^{-1/2}\cdot n^{-1})
\end{aligned}
$$

Then applying Lemma 8.2 and noticing that $\mathcal{M}(\rho_n, n; R)$ dominates all of $\rho_n^{-1/2} \cdot n^{-1}$, $(\rho_n \cdot n)^{-1}$ and $n^{-1}\log n$ completes the proof of (8.54).

Next, we focus on proving (8.56). In this proof, we shall set $\gamma = n$ in Esseen's smoothing lemma and break the integration range into three parts: $|t| \in (0, n^{\epsilon})$, $(n^{\epsilon}, n^{1/2})$ and $(n^{1/2}, n)$

LEMMA 8.3. *We have the following bounds:*

(a). *For any fixed $\epsilon > 0$, we have*

$$\int_{n^{\epsilon}}^{n} \left| \frac{\textit{Ch.f.}^{1}(G_n; t)}{t} \right| \mathrm{d}t = O(n^{-1})$$

(b). *For a small enough constant $c_\rho > 0$, if $\rho_n \leqslant c_\rho (\log n)^{-1}$, we have*

$$\int_{C_1 n^{1/2}}^{n} \left| \frac{\mathbb{E}\left[e^{\mathrm{i}t(\widetilde{T}_n + \widetilde{\Delta}_n)}\right]}{t} \right| \mathrm{d}t = O(n^{-1})$$

for an arbitrary constant $C_1 > 0$.

(c). *For a small enough constant $C_1 > 0$ and arbitrary fixed $\epsilon > 0$, we have*

$$\int_{n^{\epsilon}}^{C_1 n^{1/2}} \left| \frac{\mathbb{E}\left[e^{\mathrm{i}t(\widetilde{T}_n + \widetilde{\Delta}_n)}\right]}{t} \right| \mathrm{d}t = O(n^{-1}\log n).$$

(d). *For fixed $\epsilon > 0$ chosen such that $\epsilon \leqslant 1/7$, then we have*

$$\int_{0}^{n^{\epsilon}} \left| \frac{\mathbb{E}\left[e^{\mathrm{i}t(\widetilde{T}_n + \widetilde{\Delta}_n)}\right] - \textit{Ch.f.}(G_n; t)}{t} \right| \mathrm{d}t = O((\rho_n \cdot n)^{-1} + n^{-1}\log n).$$

PROOF OF LEMMA 8.3. First of all, we notice that between two parts $\widetilde{T}_n$ and $\widetilde{\Delta}_n$, the former is completely determined by W, and the latter follows $N(0, (\rho_n \cdot n)^{-1} \cdot \sigma_w^2)$, where $\sigma_w^2 \asymp 1$ is a U-statistic of $X_1, \ldots, X_n$. We have

$$\begin{aligned}
\mathbb{E}\left[e^{\mathrm{i}t\widetilde{T}_n} \cdot e^{\mathrm{i}t\widetilde{\Delta}_n}\right] &= \mathbb{E}\left[\mathbb{E}\left[e^{\mathrm{i}t\widetilde{T}_n} \cdot e^{\mathrm{i}t\widetilde{\Delta}_n} \middle| W\right]\right] = \mathbb{E}\left[e^{\mathrm{i}t\widetilde{T}_n} \cdot \mathbb{E}\left[e^{\mathrm{i}t\widetilde{\Delta}_n} \middle| W\right]\right] \\
&= \mathbb{E}\left[e^{\mathrm{i}t\widetilde{T}_n} \cdot e^{-(\rho_n \cdot n)^{-1}\sigma_w^2 t^2/2}\right]
\end{aligned}$$

Then we prove each of the bounds in the lemma.

(a). Notice that for each of $k = -1, 0, 1, 2, 3, \ldots$, we always have $t^k e^{-t^2/2} \leqslant C_k e^{-t^2/3}$ when $t > 1$ for universal constants $C_k > 0$ that only depend on k. From the classical literature on Hermite polynomials, we recall that function Ch.f.$(G_n; t)$ takes the form of $e^{-t^2/2}$ multiplies a polynomial of t. Therefore, for $k = -1, 0, 1, 2, 3 \ldots$

$$\int_{n^{\epsilon}}^{n} |\text{Ch.f.}(G_n; t)/t| \mathrm{d}t \leqslant (C_{-1} + \cdots + C_{d_g - 1}) \int_{n^{\epsilon}}^{\infty} e^{-t^2/3} \mathrm{d}t = O(n^{-1})$$

where $d_g :=$ degree of Ch.f.$(G_n; t)$ is a fixed finite number.

[1]Ch.f.: characteristic function. For the Edgeworth expansion function G_n that is not necessarily a valid CDF, its Ch.f. is defined to be its Fourier transform.

(b). For $|t| \ge n^{1/2}$, we have

$$\left|\mathbb{E}\left[e^{\mathbb{i}t\widetilde{T}_n}\cdot e^{-(\rho_n\cdot n)^{-1}\sigma_w^2 t^2/2}\right]\right| \leqslant \mathbb{E}\left[\left|e^{\mathbb{i}t\widetilde{T}_n}\right|\cdot\left|e^{-(\rho_n\cdot n)^{-1}\sigma_w^2 t^2/2}\right|\right]$$

$$= \mathbb{E}\left[e^{-(\rho_n\cdot n)^{-1}\sigma_w^2 t^2/2}\right] \leqslant \mathbb{E}\left[e^{-(\rho_n\cdot n)^{-1}\mathbb{E}[\sigma_w^2]/4\cdot t^2}\right] + \mathbb{P}\left(\sigma_w^2 < \mathbb{E}[\sigma_w^2]/4\right)$$

$$\leqslant e^{-C_1\cdot\rho_n^{-1}} + e^{-C_2 n} = Cn^{-C_1\cdot c_\rho^{-1}} \tag{8.57}$$

since $\rho_n^{-1} = c_\rho^{-1}\log n$, and notice that $\mathbb{P}\left(\sigma_w^2 < \mathbb{E}[\sigma_w^2]/4\right)$ diminishes exponentially fast because σ_w^2 is a U-statistic (as will be proved in the proof of part (c) below) dominated by its linear part and concentration inequalities such as Bernstein's. Then choosing $c_\rho = (4C_1)^{-1}$ finishes the proof of Lemma 8.3-(b) since

$$\int_{C_1 n^{1/2}}^{n} t^{-1}\,\mathrm{d}t = O(\log n)$$

(c). For this part of the proof, we show that σ_w^2 can be written as the sum of U-statistics thus Hoeffding's decomposition to U-statistics conveniently applies to it[2]. Then we combine this argument with the argument used in [21]. Recall that $\widehat{\Theta}_{ij} \asymp \rho_n^{-1}\cdot n^{1/2}$, and it is a U-statistic with the index set $\{1,\ldots,n\}\backslash\{i,j\}$, thus the Hoeffding's decomposition implies:

$$\widehat{\Theta}_{ij}\cdot\rho_n\cdot n^{-1/2} = \theta_{ij} + \frac{C}{n-2}\sum_{\substack{1\leqslant k\leqslant n\\ k\neq i,j}}\breve{g}_1(X_k;X_i,X_j) + \widetilde{O}_p(n^{-1}\cdot\log n) \tag{8.58}$$

where $\theta_{ij} := \mathbb{E}[\widehat{\Theta}_{ij}|X_i,X_j]\cdot\rho_n\cdot n^{-1/2}$, and we used [97] to obtain a probabilistic upper bound of the higher order terms in Hoeffding's decomposition. Then we have

$$\sigma_w^2 = \rho_n\cdot n\cdot\mathrm{Var}\left(\frac{1}{\binom{n}{2}}\sum_{1\leqslant i<j\leqslant n}\widehat{\Theta}_{ij}\eta_{ij}\middle|W\right) = \frac{\rho_n\cdot n}{\binom{n}{2}^2}\sum_{1\leqslant i<j\leqslant n}\widehat{\Theta}_{ij}^2 W_{ij}(1-W_{ij})$$

$$= \frac{\rho_n\cdot n}{\binom{n}{2}^2}\cdot\rho_n^{-2}\cdot n\cdot\sum_{1\leqslant i<j\leqslant n}\left\{\theta_{ij} + \frac{C}{n-2}\sum_{\substack{1\leqslant k\leqslant n\\ k\neq i,j}}\breve{g}_1(X_k;X_i,X_j) + \widetilde{O}_p(n^{-1}\cdot\log n)\right\}^2\cdot W_{ij}(1-W_{ij})$$

$$= \frac{\rho_n^{-1}n^2}{\binom{n}{2}^2}\sum_{1\leqslant i<j\leqslant n}\left\{\theta_{ij}^2 + \frac{2C\theta_{ij}}{n-2}\sum_{\substack{1\leqslant k\leqslant n\\ k\neq i,j}}\breve{g}_1(X_k;X_i,X_j) + \widetilde{O}_p(n^{-1}\cdot\log n)\right\}W_{ij}(1-W_{ij})$$

$$= \frac{\rho_n^{-1}\cdot n^2\sum_{1\leqslant i<j\leqslant n}\theta_{ij}^2 W_{ij}(1-W_{ij})}{\binom{n}{2}^2}$$

$$+ \frac{\rho_n^{-1}\cdot n^2\cdot 2C}{(n-2)\cdot\binom{n}{2}^2}\sum_{\substack{1\leqslant i<j\leqslant n\\ 1\leqslant k\leqslant n\\ k\neq i,j}}\breve{g}_1(X_k;X_i,X_j)W_{ij}(1-W_{ij}) + \widetilde{O}_p(n^{-1}\cdot\log n) \tag{8.59}$$

---

[2]Notice that in this part of the proof, we cannot simply bound the σ_w term away because it is dependent on any individual term in the expansion of $\widetilde{T}_n$.

where we used the fact $\left|(n-2)^{-1}\sum_{1\leqslant k\leqslant n;k\neq i,j}\breve{g}_1(X_k;X_i,X_j)\right| = \widetilde{O}_p\big((n^{-1}\log n)^{1/2}\big)$ by Bernstein inequality.

Clearly, the first term in (8.59) is a U-statistic of degree 2, where the individual term is at the order

$$\frac{\rho_n^{-1}\cdot n^2\cdot\theta_{ij}^2\cdot W_{ij}(1-W_{ij})}{\binom{n}{2}} \asymp \frac{\rho_n^{-1}\cdot n^2\cdot 1\cdot\rho_n}{n^2} \asymp 1$$

Now we focus on the second term and re-express it as a U-statistic. We have

$$\begin{aligned}
&\sum_{\substack{1\leqslant i<j\leqslant n\\ 1\leqslant k\leqslant n\\ k\neq i,j}} \theta_{ij}\breve{g}_1(X_k;X_i,X_j)W_{ij}(1-W_{ij}) = \frac{1}{2}\sum_{\substack{1\leqslant\{i,j,k\}\leqslant n\\ i\neq j,j\neq k,k\neq i}} \theta_{ij}\breve{g}_1(X_k;X_i,X_j)W_{ij}(1-W_{ij})\\
&=\frac{1}{2}\sum_{\substack{1\leqslant\{i,j,k\}\leqslant n\\ i\neq j,j\neq k,k\neq i}}\Bigg[\frac{1}{3}\Big\{\theta_{ij}\breve{g}_1(X_k;X_i,X_j)W_{ij}(1-W_{ij})\\
&\quad+\theta_{ki}\breve{g}_1(X_j;X_k,X_i)W_{ki}(1-W_{ki})+\theta_{jk}\breve{g}_1(X_i;X_j,X_k)W_{jk}(1-W_{jk})\Big\}\Bigg]\\
&=: \sum_{1\leqslant i<j<k\leqslant n}\breve{H}(X_i,X_j,X_k) \qquad (8.60)
\end{aligned}$$

where we denote

$$\begin{aligned}
\breve{H}(X_i,X_j,X_k) &:= \theta_{ij}\breve{g}_1(X_k;X_i,X_j)W_{ij}(1-W_{ij})\\
&+\theta_{ki}\breve{g}_1(X_j;X_k,X_i)W_{ki}(1-W_{ki})+\theta_{jk}\breve{g}_1(X_i;X_j,X_k)W_{jk}(1-W_{jk})
\end{aligned}$$

Clearly, $\breve{H}(X_i,X_j,X_k)$ is symmetric in X_i,X_j,X_k, and the individual term

$$\frac{\rho_n^{-1}\cdot n^2\cdot 2C\cdot\binom{n}{3}}{(n-2)\cdot\binom{n}{2}^2}\cdot\breve{H}(X_i,X_j,X_k) \asymp \frac{\rho_n^{-1}\cdot n^2\cdot n^3}{n^5}\cdot\rho_n \asymp 1$$

So the second term on the RHS of (8.59) is a U-statistic of degree 3. Therefore, σ_w^2 can be re-expressed as Hoeffding's decomposition for U-statistics as follows

$$\sigma_w^2 = \mathbb{E}[\sigma_w^2] + \frac{1}{n}\sum_{i=1}^n g_{\sigma;1}(X_i) + \widetilde{O}_p(n^{-1}\cdot\log n) \qquad (8.61)$$

where we again applied [97] to derive the probabilistic upper bound for the higher order terms in Hoeffding's decomposition.

Now, we are ready to upper bound the characteristic function for $n^\epsilon \preceq |t| \preceq n^{1/2}$

$$\begin{aligned}
&\left|\mathbb{E}\left[e^{\mathrm{i}t\widetilde{T}_n}\cdot e^{-(\rho_n\cdot n)^{-1}\sigma_w^2t^2/2}\right]\right|\\
&\leqslant\left|\mathbb{E}\left[e^{\mathrm{i}t\widetilde{T}_n}\cdot e^{-(\rho_n\cdot n)^{-1}t^2/2\cdot\left\{\mathbb{E}[\sigma_w^2]+\frac{1}{n}\sum_{i=1}^n g_{\sigma;1}(X_i)+\widetilde{O}_p(n^{-1}\cdot\log n)\right\}}\right]\right|\\
&=\left|\mathbb{E}\left[e^{\mathrm{i}t\widetilde{T}_n}\cdot e^{-(\rho_n\cdot n)^{-1}t^2/2\cdot\left\{\mathbb{E}[\sigma_w^2]+\frac{1}{n}\sum_{i=1}^n g_{\sigma;1}(X_i)\right\}}\cdot\left(1+\widetilde{O}_p\left(\rho_n^{-1}\cdot n^{-2}\log n\cdot t^2\right)\right)\right]\right| \qquad (8.62)
\end{aligned}$$

where in the last line, we used the fact that $|e^z - 1| = O(|z|)$ for all universally bounded $z \in \mathbb{C}$ (here $|t| = O(n^{1/2})$ and by assumption $\rho_n \cdot \log n = O(1)$). Then since

$$\int_{n^\epsilon}^{n^{1/2}} \frac{\rho_n^{-1} \cdot n^{-2} \cdot \log n \cdot t^2}{t} \mathrm{d}t \asymp (\rho_n \cdot n)^{-1} \cdot \log n \tag{8.63}$$

we know that this $\widetilde{O}_p(\rho_n^{-1} \cdot n^{-2} \cdot \log n \cdot t^2)$ term can be ignored in (8.62). Continuing (8.62), we have

$$\text{RHS of (8.62)} \leqslant e^{-(\rho_n \cdot n)^{-1} t^2/2 \cdot \mathbb{E}[\sigma_w^2]} \cdot \left|\mathbb{E}\left[e^{\mathbb{i}t\widetilde{T}_n} \cdot e^{-\rho_n^{-1} \cdot n^{-2} \cdot \sum_{i=1}^n g_{\sigma;1}(X_i) \cdot t^2}\right]\right|$$

We are going to show that $\widetilde{T}_n$ can be expressed as a U-statistic of degree 2 plus an $\widetilde{O}_p(n^{-1}\log^{3/2} n)$ remainder term, which can be ignored. Indeed,

$$\begin{aligned}
\widetilde{T}_n =& U_n^{\#} + \Delta_n - \frac{1}{2} \cdot U_n^{\#} \cdot \delta_n \\
=& \frac{1}{\sqrt{n}\xi_1} \sum_{i=1}^n g_1(X_i) + \frac{r-1}{\sqrt{n}(n-1)\xi_1} \sum_{1 \leqslant i < j \leqslant n} g_2(X_i, X_j) \\
&+ \frac{1}{n^{3/2}\xi_1} \sum_{i=1}^n g_1(X_i) \sum_{j=1}^n \frac{g_1^2(X_j) - \xi_1^2}{\xi_1^2} + \widetilde{O}_p(n^{-1} \cdot \log^{3/2} n).
\end{aligned}$$

Since $n^{-3/2} \sum_{i=1}^n g_1(X_i)\big(g_1(X_i)^2 - \xi_1^2\big)/\xi_1^3 = \widetilde{O}_p(n^{-1} \cdot \log^{1/2} n)$, we can write

$$\begin{aligned}
\widetilde{T}_n =& \frac{1}{\sqrt{n}\xi_1} \sum_{i=1}^n g_1(X_i) + \frac{r-1}{\sqrt{n}(n-1)\xi_1} \sum_{1 \leqslant i < j \leqslant n} g_2(X_i, X_j) \\
&+ \frac{1}{n^{3/2}\xi_1} \sum_{1 \leqslant i < j \leqslant n}^{n} \frac{g_1(X_i)\big(g_1^2(X_j) - \xi_1^2\big) + g_1(X_j)\big(g_1^2(X_i) - \xi_1^2\big)}{\xi_1^2} + \widetilde{O}_p(n^{-1} \cdot \log^{3/2} n) \\
&=: \frac{1}{\sqrt{n}\xi_1} \sum_{i=1}^n g_1(X_i) + \frac{r-1}{\sqrt{n}(n-1)} \sum_{1 \leqslant i < j \leqslant n} \widetilde{g}_2(X_i, X_j) + \widetilde{O}_p(n^{-1} \cdot \log^{3/2} n)
\end{aligned}$$

which therefore is expressed as a U-statistic of degree 2 plus an $\widetilde{O}_p(n^{-1} \cdot \log^{3/2} n)$ term, where $\mathbb{E}[\widetilde{g}_2(X_i, X_j)] = 0$ and $\mathbb{E}[\widetilde{g}_2^2(X_i, X_j)] = O(1)$. To prove the claimed bound, we can choose a positive integer m (depending on t) and write

$$\sum_{1 \leqslant i < j \leqslant n} \widetilde{g}_2(X_i, X_j) = \sum_{i=1}^m \sum_{j=i+1}^n \widetilde{g}_2(X_i, X_j) + \sum_{i=m+1}^{n-1} \sum_{j=i+1}^n \widetilde{g}_2(X_i, X_j)$$

Then the arguments of [21, eq. (2.17)-(2.20)] can be applied here. Notice that this part of the proof of [21] does not require non-lattice assumption, but all it requires on the behavior of $|\mathbb{E}[e^{\mathbb{i}tg_1(X_i)/(\sqrt{n}\cdot\xi_1)}]|$ is its closeness to 1 for $t/\sqrt{n} \approx 0$. Indeed, for $n\rho_n \gg 1$ and $t \leqslant c_1 n^{1/2}$ with small $c_1 > 0$,

$$\begin{aligned}
&\big|\mathbb{E}e^{\mathbb{i}tg_1(X_i)/(\sqrt{n}\cdot\xi_1) - \rho_n^{-1}n^{-2}t^2/2g_{\sigma,1}(X_i)}\big| \\
&\leqslant \left|\mathbb{E}\left(1 + \frac{1}{2}\Big(\frac{\mathbb{i}tg_1(X_i)}{\sqrt{n}\xi_1} - \frac{t^2 g_{\sigma,1}(X_i)}{2\rho_n n^2}\Big)^2\right)\right| + O\left(\mathbb{E}\Big|\frac{\mathbb{i}tg_1(X_i)}{\sqrt{n}\xi_1} - \frac{t^2 g_{\sigma,1}(X_i)}{2\rho_n n^2}\Big|^3\right) \\
&\leqslant 1 - \frac{t^2}{3n} \leqslant \exp\Big\{-\frac{t^2}{3n}\Big\}.
\end{aligned}$$

The proof of Lemma 8.3-(c) is therefore completed after applying the arguments of [21, eq. (2.17)-(2.20)].

(d). Finally, in this part, we calculate the expansion of $\mathbb{E}\left[e^{\mathbb{i}t\widetilde{T}_n}\right]$ and derive the Edgeworth expansion for $|t| \leqslant n^{\epsilon}$ for a small enough fixed ϵ. The main portion of the proof for this part, i.e., our calculations in (8.69), (8.70), (8.73) and (8.74) that we are going to present, follow the roadmap in classical literature on Edgeworth expansion for noiseless U-statistics, laid out by [21, 71, 90, 96]. Our $\widetilde{T}_n$ is different from their studentization/standardization forms by using a different rescaler, so this part is not a direct corollary of their results. Despite the resulting differences is non-essential, we nonetheless present the full calculation steps for completeness and for the convenience of the readers.

To start, we have

$$\begin{aligned}
&\mathbb{E}\Big[e^{\mathbb{i}t\widetilde{T}_n} \cdot e^{-(\rho_n \cdot n)^{-1}\sigma_w^2 t^2/2}\Big] \\
&= \mathbb{E}\left[e^{\mathbb{i}t\widetilde{T}_n} \cdot \left\{1 - \frac{\sigma_w^2 t^2}{2\rho_n \cdot n} + \frac{\sigma_w^4 t^4}{8\rho_n^2 \cdot n^2} + O\left(\frac{\sigma_w^6 t^6}{\rho_n^3 \cdot n^3}\right)\right\}\right] \tag{8.64}
\end{aligned}$$

as long as $n\rho_n = \omega(n^{2\epsilon})$. We first bound the remainder, we have $\int_0^{n^\epsilon} (\sigma_w^6 t^6)(\rho_n^3 n^3) \cdot t^{-1}\mathrm{d}t \asymp n^{6\epsilon} \cdot (\rho_n \cdot n)^{-3}$. Since the assumption of Theorem 3.1 implies that $\rho_n = \omega(n^{-1/2})$ in any case, so setting $\epsilon \leqslant 1/13$ yields $n^{6\epsilon} \cdot (\rho_n \cdot n)^{-3} = O(n^{-1})$. We have

$$\begin{aligned}
e^{\mathbb{i}t\widetilde{T}_n} &= e^{\mathbb{i}t\left(U_n^{\#} + \Delta_n - \frac{1}{2}U_n^{\#}\cdot\delta_n\right)} \\
&= e^{\mathbb{i}tU_n^{\#}}\left\{1 + \left(\Delta_n - \frac{1}{2}U_n^{\#}\cdot\delta_n\right)\mathbb{i}t - \frac{1}{2}\cdot\left(\Delta_n - \frac{1}{2}U_n^{\#}\cdot\delta_n\right)^2 t^2\right\} \\
&\quad + \widetilde{O}_p\left(\left|\Delta_n - \frac{1}{2}U_n^{\#}\cdot\delta_n\right|^3 t^3\right) \tag{8.65}
\end{aligned}$$

To bound the remainder term, notice that $\left|1 - \sigma_w^2 t^2/(\rho_n \cdot n)\right| \leqslant 1$ for $|t| \leqslant n^{\epsilon}$, where we recall that Theorem 3.1 we are proving here always assumes $\rho_n = \omega(n^{-1/2})$ in all cases. Then, setting $\epsilon \leqslant 1/7$ together with the fact $U_n^{\#} = \widetilde{O}_p(\log^{1/2} n)$, $\Delta_n = \widetilde{O}_p(n^{-1/2}\log n)$, $\delta_n = \widetilde{O}_p(n^{-1/2}\log^{1/2} n)$, by Bernstein's inequality and [97], we have

$$\begin{aligned}
\int_0^{n^\epsilon} \left|\Delta_n - \frac{1}{2}U_n^{\#}\cdot\delta_n\right|^3 t^3 \cdot \frac{1}{t}\mathrm{d}t &= \widetilde{O}_p\left(n^{-3/2}\cdot n^{3\epsilon}\log^{3/2} n\right) \\
&= \widetilde{O}_p(n^{-15/14}\log^{3/2} n) = \widetilde{O}_p(n^{-1}) \tag{8.66}
\end{aligned}$$

and this remainder term can also be ignored. Now we deal with the main part of the terms. Set $\varphi_n(t) := \mathbb{E}\left[e^{\mathbb{i}t\cdot\frac{g_1(X_1)}{\sqrt{n}\cdot\xi_1}}\right]$. Then by Section VI, Lemma 4 of [108], we have

$$\varphi_n^n(t) = e^{-t^2/2}\left(1 - n^{-1/2}\cdot\frac{\mathbb{i}\mathbb{E}[g_1^3(X_1)]t^3}{6\xi_1^3}\right) + O\left(n^{-1}\cdot\log n\cdot P_0(t)e^{-t^2/4}\right) \tag{8.67}$$

$$\varphi_n^{n-k}(t) = \varphi_n^n(t) + O\left(n^{-1}\cdot\log n\cdot P_k(t)e^{-t^2/4}\right)$$

for any fixed $k = 0, 1, 2, 3$, where $P_0(t), \ldots, P_k(t)$ are fixed polynomials of t and each of them can be divided by t. Here, we first focus on $\mathbb{E}[e^{\mathbb{i}t\widetilde{T}_n}]$, and then handle $\mathbb{E}[e^{\mathbb{i}t\widetilde{T}_n}\cdot\sigma_w^2 t^2/(\rho_n\cdot n)]$. For $\mathbb{E}[e^{\mathbb{i}t\widetilde{T}_n}]$, by ignoring the small term in (8.65), we have

$$\mathbb{E}\left[e^{\mathbb{i}t\widetilde{T}_n}\right] = \mathbb{E}\left[e^{\mathbb{i}tU_n^{\#}}\left\{1 + \mathbb{i}t\left(\Delta_n - \frac{1}{2}U_n^{\#}\cdot\delta_n\right) - \frac{t^2}{2}\left(\Delta_n - \frac{1}{2}U_n^{\#}\cdot\delta_n\right)^2\right\}\right] \tag{8.68}$$

Now we inspect each term on the RHS of (8.68). For $\mathbb{E}[e^{\mathbb{i}tU_n^{\#}}]$ we use (8.67) and obtain $\mathbb{E}[\mathbb{i}tU_n^{\#}] = \varphi_n^n(t)$. For the next term, recall that $\mathbb{E}[g_2(X_1,X_2)] = 0$ and $\mathbb{E}[g_1^k(X_1)g_2(X_1,X_2)] = 0$ for all $k \in \mathbb{N}$. We have

$$\mathbb{E}\left[e^{\mathbb{i}tU_n^{\#}} \cdot \mathbb{i}t\Delta_n\right] = \mathbb{E}\left[e^{\mathbb{i}tU_n^{\#}} \cdot \mathbb{i}t \cdot \frac{r-1}{\sqrt{n}(n-1)} \sum_{1\leqslant i<j\leqslant n} \frac{g_2(X_i,X_j)}{\xi_1}\right]$$

$$=\frac{\mathbb{i}t(r-1)}{\sqrt{n}(n-1)} \cdot \binom{n}{2} \cdot \varphi_n^{n-2}(t) \cdot \mathbb{E}\left[e^{\mathbb{i}t\frac{g_1(X_1)+g_1(X_2)}{\sqrt{n}\xi_1}} \cdot \frac{g_2(X_1,X_2)}{\xi_1}\right]$$

$$=\frac{\mathbb{i}t(r-1)\sqrt{n}}{2} \cdot \varphi_n^{n-2}(t) \cdot \mathbb{E}\left[\frac{g_2(X_1,X_2)}{\xi_1} + \frac{\mathbb{i}t(g_1(X_1)+g_1(X_2))g_2(X_1,X_2)}{\sqrt{n}\cdot\xi_1^2}\right.$$

$$\left. - \frac{t^2\left\{g_1^2(X_1)+2g_1(X_1)g_1(X_2)+g_1^2(X_2)\right\}\cdot g_2(X_1,X_2)}{2n\xi_1^3} + \right] + O\left(n^{-1}\cdot e^{-t^2/4}\cdot \text{Poly}(t)\right)$$

$$=\frac{\mathbb{i}t(r-1)\sqrt{n}}{2} \cdot \varphi_n^{n-2}(t) \cdot \mathbb{E}\left[\frac{g_2(X_1,X_2)}{\xi_1} + \frac{2\mathbb{i}tg_1(X_1)g_2(X_1,X_2)}{\sqrt{n}\cdot\xi_1^2}\right.$$

$$\left. - \frac{t^2\left\{g_1^2(X_1)+g_1(X_1)g_1(X_2)\right\}\cdot g_2(X_1,X_2)}{n\xi_1^3}\right] + O\left(n^{-1}\cdot e^{-t^2/4}\cdot \text{Poly}(t)\right)$$

$$=\frac{-\mathbb{i}t^3(r-1)}{2\sqrt{n}\cdot\xi_1^3} \cdot \varphi_n^{n-2}(t) \cdot \mathbb{E}\left[g_1(X_1)g_1(X_2)\cdot g_2(X_1,X_2)\right] + O\left(n^{-1}\cdot e^{-t^2/4}t\cdot \text{Poly}(t)\right)$$

(8.69)

$$=e^{-t^2/2} \cdot \frac{-\mathbb{i}t^3(r-1)}{2\sqrt{n}\cdot\xi^3} \cdot \mathbb{E}\left[g_1(X_1)g_1(X_2)g_2(X_1,X_2)\right] + O\left(n^{-1}\cdot e^{-t^2/4}t\cdot \text{Poly}(t)\right)$$

We use the approximation to δ_n given by Lemma 3.1-(d). When we use it here, we may ignore any $\widetilde{O}_p(n^{-1}\log n)$ remainder term, which is justified by Lemma 8.2 in the real domain, not the frequency domain that characteristic function works with. We thus have

$$\mathbb{E}\left[e^{\mathbb{i}tU_n^{\#}} \cdot \mathbb{i}t\left(-\frac{1}{2}U_n^{\#}\cdot\delta_n\right)\right] = -\frac{1}{2}\mathbb{i}t \cdot \mathbb{E}\left[e^{\mathbb{i}t\frac{\sum_{i=1}^n g_1(X_i)}{\sqrt{n}\cdot\xi_1}} \cdot \left\{\frac{\sum_{i=1}^n g_1(X_i)}{\sqrt{n}\cdot\xi_1}\right\}\cdot\right.$$

(8.70)

$$\left.\cdot\left(\frac{\sum_{j=1}^n\left\{g_1^2(X_j)-\xi_1^2\right\}}{n\xi_1^2} + \frac{2(r-1)\sum_{i=1}^n\sum_{j\neq i}g_1(X_i)g_2(X_i,X_j)}{n(n-1)\xi_1^2}\right) + \widetilde{O}_p(n^{-1}\cdot\log^{3/2}n)\right]$$

We consider the expression into two parts by the two terms inside the parenthesis on the RHS of the equation, and inspect them respectively. Ignoring the $\widetilde{O}_p(n^{-1}\cdot\text{Polylog}(n))$ remainder, for the first part, we have

$$-\frac{1}{2}\mathbb{i}t \cdot \mathbb{E}\left[e^{\mathbb{i}t\frac{\sum_{i=1}^n g_1(X_i)}{\sqrt{n}\cdot\xi_1}} \cdot \left\{\frac{\sum_{i=1}^n g_1(X_i)}{\sqrt{n}\cdot\xi_1}\right\}\cdot\left(\frac{\sum_{j=1}^n\left\{g_1^2(X_j)-\xi_1^2\right\}}{n\xi_1^2}\right)\right]$$

(8.71)

$$=-\frac{1}{2}\mathbb{i}t \cdot \mathbb{E}\left[e^{\mathbb{i}t\frac{\sum_{i=1}^n g_1(X_i)}{\sqrt{n}\cdot\xi_1}} \cdot \left\{\sum_{i=1}^n \frac{g_1(X_i)\left(g_1^2(X_i)-\xi_1^2\right)}{n\sqrt{n}\cdot\xi_1^3} + \sum_{\substack{i,j\in\{1,\dots,n\}\\ i\neq j}} \frac{g_1(X_i)\left(g_1^2(X_j)-\xi_1^2\right)}{n\sqrt{n}\cdot\xi_1^3}\right\}\right]$$

Further breaking the RHS down and handle the two summations in the fancy bracket separately, we have

$$
\begin{aligned}
&-\frac{1}{2}\mathbb{i}t\cdot\mathbb{E}\left[e^{\mathbb{i}t\frac{\sum_{i=1}^n g_1(X_i)}{\sqrt{n}\cdot\xi_1}}\cdot\left\{\sum_{i=1}^n\frac{g_1(X_i)\left(g_1^2(X_i)-\xi_1^2\right)}{n\sqrt{n}\cdot\xi_1^3}\right\}\right]\\
&=-\frac{1}{2}\mathbb{i}t\cdot\varphi_n^{n-1}(t)\cdot n\cdot\mathbb{E}\left[\left\{1+\frac{\mathbb{i}t\cdot g_1(X_1)}{\sqrt{n}\cdot\xi_1}\right\}\cdot\left\{\frac{g_1(X_1)\left(g_1^2(X_1)-\xi_1^2\right)}{n\sqrt{n}\cdot\xi_1^3}\right\}\right]\\
&\quad+O\left(n^{-1}\cdot e^{-t^2/4}t^2\cdot\text{Poly(t)}\right)\\
&=-\frac{1}{2}\cdot\frac{\mathbb{i}t\varphi_n^{n-1}(t)}{\sqrt{n}\cdot\xi_1^3}\cdot\mathbb{E}\left[g_1^3(X_1)\right]+O\left(n^{-1}\cdot e^{-t^2/4}t^2\cdot\text{Poly(t)}\right)
\end{aligned}
\tag{8.72}
$$

and

$$
\begin{aligned}
&-\frac{1}{2}\mathbb{i}t\cdot\mathbb{E}\left[e^{\mathbb{i}t\frac{\sum_{i=1}^n g_1(X_i)}{\sqrt{n}\cdot\xi_1}}\cdot\left\{\sum_{\substack{i,j\in\{1,\dots,n\}\\ i\neq j}}\frac{g_1(X_i)\left(g_1^2(X_j)-\xi_1^2\right)}{n\sqrt{n}\cdot\xi_1^3}\right\}\right]\\
&=-\frac{1}{2}\mathbb{i}t\cdot\varphi_n^{n-2}(t)\cdot n(n-1)\cdot\mathbb{E}\left[\left\{1+\mathbb{i}t\cdot\frac{g_1(X_1)}{\sqrt{n}\cdot\xi_1}-\frac{t^2g_1^2(X_1)}{2n\xi_1^2}\right\}\right.\\
&\left.\left\{1+\mathbb{i}t\cdot\frac{g_1(X_2)}{\sqrt{n}\cdot\xi_1}-\frac{t^2g_1^2(X_2)}{2n\xi_1^2}\right\}\cdot\left\{\frac{g_1(X_1)\left(g_1^2(X_2)-\xi_1^2\right)}{n\sqrt{n}\cdot\xi_1^3}\right\}\right]+O\left(n^{-1}\cdot e^{-t^2/4}t^2\cdot\text{Poly(t)}\right)\\
&=\frac{1}{2}\mathbb{i}t^3\cdot\varphi_n^{n-2}(t)\cdot n(n-1)\cdot\mathbb{E}\left[\frac{g_1^2(X_1)g_1(X_2)\left\{g_1^2(X_2)-\xi_1^2\right\}}{n^2\sqrt{n}\cdot\xi_1^5}\right]+O\left(n^{-1}\cdot e^{-t^2/4}t^2\cdot\text{Poly(t)}\right)\\
&=\frac{1}{2}\frac{\mathbb{i}t^3\varphi_n^{n-2}(t)}{\sqrt{n}\cdot\xi_1^3}\cdot\mathbb{E}\left[g_1^3(X_1)\right]+O\left(n^{-1}\cdot e^{-t^2/4}t^2\cdot\text{Poly(t)}\right)
\end{aligned}
\tag{8.73}
$$

Now we calculate Part 2 of the RHS of (8.70). We have

$$
\begin{aligned}
&-\frac{1}{2}\mathbb{i}t\cdot\mathbb{E}\left[e^{\mathbb{i}t\frac{\sum_{i=1}^n g_1(X_i)}{\sqrt{n}\cdot\xi_1}}\cdot\left\{\frac{\sum_{i=1}^n g_1(X_i)}{\sqrt{n}\cdot\xi_1}\right\}\cdot\left(\frac{2(r-1)\sum_{i=1}^n\sum_{j\neq i}g_1(X_i)g_2(X_i,X_j)}{n(n-1)\xi_1^2}\right)\right]\\
&=-\frac{(r-1)\mathbb{i}t}{\xi_1^2}\cdot\mathbb{E}\left[e^{\mathbb{i}t\frac{\sum_{i=1}^n g_1(X_i)}{\sqrt{n}\cdot\xi_1}}\cdot\left\{\frac{g_1(X_1)+g_1(X_2)}{\sqrt{n}\cdot\xi_1}\cdot g_1(X_1)g_2(X_1,X_2)\right\}\right]\\
&\quad-\frac{(r-1)\mathbb{i}t}{\xi_1^2}(n-2)\cdot\mathbb{E}\left[e^{\mathbb{i}t\frac{\sum_{i=1}^n g_1(X_i)}{\sqrt{n}\cdot\xi_1}}\cdot\left\{\frac{g_1(X_3)}{\sqrt{n}\cdot\xi_1}\cdot g_1(X_1)g_2(X_1,X_2)\right\}\right]\\
&=-\frac{(r-1)\mathbb{i}t}{\sqrt{n}\cdot\xi_1^3}\cdot\varphi_n^{n-2}(t)\cdot\mathbb{E}\left[g_1(X_1)g_1(X_2)g_2(X_1,X_2)\right]+O\left(n^{-1}\cdot e^{-t^2/4}t^2\cdot\text{Poly(t)}\right)\\
&\quad-\frac{(r-1)\mathbb{i}t}{\sqrt{n}\cdot\xi_1^3}(n-2)\cdot\varphi_n^{n-3}(t)\cdot\mathbb{E}\left[e^{\mathbb{i}t\frac{g_1(X_1)}{\sqrt{n}\cdot\xi_1}}\cdot\left\{1+\frac{\mathbb{i}tg_1(X_2)}{\sqrt{n}\cdot\xi_1}-\frac{t^2g_1^2(X_2)}{2n\xi_1^2}\right\}\right.\\
&\quad\left.\cdot\left\{1+\frac{\mathbb{i}tg_1(X_3)}{\sqrt{n}\cdot\xi_1}-\frac{t^2g_1^2(X_3)}{2n\xi_1^2}\right\}\cdot g_1(X_1)g_2(X_1,X_2)g_1(X_3)\right]
\end{aligned}
$$

$$= -\frac{(r-1)\mathbb{i}t}{\sqrt{n}\cdot\xi_1^3}\cdot\varphi_n^{n-2}(t)\cdot\mathbb{E}\left[g_1(X_1)g_1(X_2)g_2(X_1,X_2)\right]+O\left(n^{-1}\cdot e^{-t^2/4}t^2\cdot\text{Poly(t)}\right)$$

$$-\frac{(r-1)\mathbb{i}t}{\sqrt{n}\cdot\xi_1^3}(n-2)\cdot\varphi_n^{n-3}(t)\cdot\mathbb{E}\left[\frac{-t^2}{n\xi_1^2}\cdot g_1(X_1)g_1(X_2)g_2(X_1,X_2)g_1^2(X_3)\right]$$

$$=\frac{(r-1)\mathbb{i}(t^3-t)}{\sqrt{n}\cdot\xi_1^3}\cdot e^{-t^2/2}\cdot\mathbb{E}\left[g_1(X_1)g_1(X_2)g_2(X_1,X_2)\right]+O\left(n^{-1}\cdot e^{-t^2/4}t^2\cdot\text{Poly(t)}\right) \tag{8.74}$$

Collecting terms (8.69), (8.73) and (8.74), we have

$$\begin{aligned}
&\mathbb{E}\left[e^{\mathbb{i}t\left(U_n^{\#}+\widetilde{\Delta}_n+\Delta_n-\frac{1}{2}U_n^{\#}\delta_n\right)}\right]\\
=&e^{-t^2/2}\cdot\left\{1-\left(\frac{\mathbb{E}\left[g_1^3(X_1)\right]}{2}+(r-1)\mathbb{E}\left[g_1(X_1)g_1(X_2)g_2(X_1,X_2)\right]\right)\cdot\frac{\mathbb{i}t}{\sqrt{n}\cdot\xi_1^3}\right.\\
&\left.+\left(\frac{\mathbb{E}\left[g_1^3(X_1)\right]}{3}+\frac{(r-1)}{2}\mathbb{E}\left[g_1(X_1)g_1(X_2)g_2(X_1,X_2)\right]\right)\cdot\frac{\mathbb{i}t^3}{\sqrt{n}\cdot\xi_1^3}\right\}\\
&+O\left(n^{-1}\log n\cdot e^{-t^2/4}\cdot\text{Poly}(t)\right)
\end{aligned}\tag{8.75}$$

The remainder term is clearly ignorable if plugged into the Esseen's smoothing lemma. It only remains to deal with the $\sigma_w^2t^2/(\rho_n\cdot n)$ term and the $\sigma_w^4t^4/(\rho_n^2\cdot n^2)$ term in (8.64). By (8.61), we have

$$\begin{aligned}
&\mathbb{E}\left[e^{\mathbb{i}t\widetilde{T}_n}\cdot\frac{\sigma_w^2t^2}{\rho_n\cdot n}\right]\\
=&\left[e^{\mathbb{i}t\widetilde{T}_n}\left(\mathbb{E}[\sigma_w^2]+\frac{1}{n}\sum_{i=1}^n g_{\sigma;1}(X_i)+\widetilde{O}_p(n^{-1}\cdot\log n)\right)\right]\cdot\frac{t^2}{\rho_n\cdot n}\\
=&\mathbb{E}\left[e^{\mathbb{i}t\widetilde{T}_n}\right]\cdot\frac{\mathbb{E}[\sigma_w^2]t^2}{\rho_n\cdot n}+\mathbb{E}\left[e^{\mathbb{i}t\widetilde{T}_n}\cdot g_{\sigma;1}(X_1)\right]\cdot\frac{t^2}{\rho_n\cdot n}+O\left(\frac{t^2\log n}{\rho_n\cdot n^2}\right)
\end{aligned}$$

Now we discuss the three terms on the RHS. Term 1:

$$\begin{aligned}
\int_0^{n^\epsilon}\left|\mathbb{E}\left[e^{\mathbb{i}t\widetilde{T}_n}\right]\cdot\frac{\mathbb{E}[\sigma_w^2]t^2}{\rho_n\cdot n}\cdot\frac{1}{t}\right|\mathrm{d}t&=\int_0^{n^\epsilon}O\left(e^{-t^2/4}\cdot\text{Poly}(t)\right)\cdot(\rho_n\cdot n)^{-1}\\
&=O\left((\rho_n\cdot n)^{-1}\right)
\end{aligned}$$

Term 2: by mimicking the derivations in our (8.72) and also referring to (2.11) in [21], we see that

$$\left|\mathbb{E}\left[e^{\mathbb{i}t\widetilde{T}_n}\cdot g_{\sigma;1}(X_1)\right]\right|=O\left(e^{-t^2/4}\cdot\text{Poly}(t)\right)$$

Therefore, it can be bounded in exactly the same way as term 1.

For term 3, we have

$$\int_0^{n^\epsilon}\frac{t^2}{\rho_n\cdot n}\cdot\frac{1}{t}\mathrm{d}t=(\rho_n\cdot n)^{-1}\cdot n^{2\epsilon-1}\leqslant(\rho_n\cdot n)^{-1}$$

where recall that $\epsilon<1/2$. The $\sigma_w^4t^4/(\rho_n^2\cdot n^2)$ term can be bounded exactly similarly and we omit the proof here. This finishes the proof of Lemma 8.3-(d).

Now we return to the proof of Theorem 3.1. Plugging the results of Lemma 8.3 back into Lemma 8.1 completes the proof of Theorem 3.1 with the assumption $\rho_n = O((\log n)^{-1})$.

If Cramer's condition holds instead of the upper bound on ρ_n, then the derivation steps in (2.21)–(2.22) in [21] can be reproduced, where their t_N can be understood as n^{r_0} for any fixed $r_0 \in (0,1)$. It would suffice for our purpose to use any $r_0 \in (1/2, 1)$. Notice that their "r" has different meaning than ours. This extends the integrative range that our Lemma 8.3-(c) holds valid from the original range $(n^{\epsilon}, C_1 \cdot n^{1/2})$ to (n^{ϵ}, n^{r_0}), and we only need to prove Lemma 8.3-(b) on the integrative range (n^{r_0}, n) instead of $(C_1 \cdot n^{1/2}, n)$. Then our proof of Lemma 8.3-(b) can be revised into

$$\begin{aligned}
&\left|\mathbb{E}\left[e^{\mathbb{i}t\widetilde{T}_n} \cdot e^{-(\rho_n \cdot n)^{-1}\sigma_w^2 t^2/2}\right]\right| \leqslant \mathbb{E}\left[\left|e^{\mathbb{i}t\widetilde{T}_n}\right| \cdot \left|e^{-(\rho_n \cdot n)^{-1}\sigma_w^2 t^2/2}\right|\right] \\
&=\mathbb{E}\left[e^{-n^{-1}\sigma_w^2 t^2/2}\right] \leqslant \mathbb{E}\left[e^{-n^{2r_0-1}\cdot\mathbb{E}[\sigma_w^2]/4}\right] + \mathbb{P}\left(\sigma_w^2 < \mathbb{E}[\sigma_w^2]/4\right) \\
&\leqslant e^{-C_1 \cdot n^{2r_0-1}} + e^{-C_2 n} \prec n^{-2} \qquad (8.76)
\end{aligned}$$

where in the second line we replaced ρ_n by 1 to majorize.

□

8.4. *Proof of Theorem 3.2.* It is easy to verify that

$$\widetilde{O}_p(p_n)\widetilde{O}_p(q_n) = \widetilde{O}_p(p_n q_n), \quad \text{and} \quad \widetilde{O}_p(p_n) + \widetilde{O}_p(q_n) = \widetilde{O}_p(p_n + q_n) \tag{8.77}$$

We also easily have $O(p_n)\widetilde{O}_p(q_n) = \widetilde{O}_p(p_n q_n)$ since $O(\cdot)$ implies $\widetilde{O}_p(\cdot)$, but it is not guaranteed that $O_p(p_n)\widetilde{O}_p(q_n) = \widetilde{O}_p(p_n q_n)$ if the distribution of $O_p(p_n)$ is heavy tailed. The presence of edge-wise observational errors introduces extra technical complications to the proof of Theorem 3.2 beyond the analysis for empirical Edgeworth expansions for noiseless U-statistics such as [71, 96] and [110]. We shall carefully address this. By the proofs of Lemma 3.1-(c) and (d), and recall that $\widehat{\xi}_1^2 = n\widehat{S}^2/r$ and $\xi_1^2 = n\sigma_n^2/r$, we have

$$\frac{(\widehat{\xi}_1 + \xi_1)(\widehat{\xi}_1 - \xi_1)}{\rho_n^{2s}} \asymp \frac{\widehat{\xi}_1^2 - \xi_1^2}{n\sigma_n^2} \asymp \delta_n + \widehat{\delta}_n = \widetilde{O}_p(n^{-1/2}\log^{1/2} n)$$

Then noticing that $\widehat{\xi}_1/\xi_1 = 1 + \widetilde{O}_p(1)$ and thus $\widehat{\xi}_1 \asymp \xi_1 \asymp \rho_n^s$ with probability at least $1 - O(n^{-1})$, we have $\widehat{\xi}_1 - \xi_1 = \widetilde{O}_p(\rho_n^s \cdot n^{-1/2}\log^{1/2} n)$. Therefore

$$\widehat{\xi}_1^3 - \xi_1^3 = \widetilde{O}_p(\rho_n^{3s} \cdot n^{-1/2}\log^{1/2} n)$$

implying that

$$\left|\frac{1}{\sqrt{n}\xi_1^3} - \frac{1}{\sqrt{n}\widehat{\xi}_1^3}\right| = \widetilde{O}_p\Big(\frac{\log^{1/2} n}{\rho_n^{3s}\cdot n}\Big).$$

Recall that $\|F_{\widehat{T}_n}(x) - G_n(x)\|_\infty = O\left(\mathcal{M}(\rho_n, n; R)\right)$, where

$$\begin{aligned}
G_n(x) = \Phi(x) + \frac{\varphi(x)}{\sqrt{n}\cdot\xi_1^3}\cdot\Big\{\Big(\frac{x^2}{3} + \frac{1}{6}\Big)\mathbb{E}[g_1^3(X_1)] \\
+ \frac{r-1}{2}(x^2+1)\mathbb{E}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)]\Big\}.
\end{aligned}$$

As a result, in order to prove $\|\widehat{G}_n(x) - G_n(x)\|_\infty = \widetilde{O}_p\left(\mathcal{M}(\rho_n, n; R)\right)$, it suffices to show that

$$\max\Big\{\big|\widehat{\mathbb{E}}g_1^3(X_1) - \mathbb{E}g_1^3(X_1)\big|$$

$$
, \left|\widehat{\mathbb{E}}\left[g_1(X_1)g_1(X_2)g_2(X_1,X_2)\right] - \mathbb{E}\left[g_1(X_1)g_1(X_2)g_2(X_1,X_2)\right]\right|\Big\}
$$

$$
= \begin{cases} \widetilde{O}_p(\rho_n^{3s-1}\cdot n^{-1/2}\log^{1/2} n) & \text{if } R \text{ is acyclic} \\ \widetilde{O}_p(\rho_n^{3s-2/r}\cdot n^{-1/2}\log^{1/2} n) & \text{if } R \text{ is cyclic} \end{cases}
$$

where we used the fact that $\sup_{x\in\mathbb{R}}|x|^3\varphi(x) = O(1)$. We will show that the empirical moments $\widehat{\mathbb{E}}\left[g_1^3(X_1)\right]$ and $\widehat{\mathbb{E}}\left[g_1(X_1)g_1(X_2)g_2(X_1,X_2)\right]$ converge to $\mathbb{E}\left[g_1^3(X_1)\right]$ and $\mathbb{E}\left[g_1(X_1)g_1(X_2)g_2(X_1,X_2)\right]$, respectively, at rates no slower than $\widetilde{O}_p(\rho_n^{3s-0.5}\cdot n^{-1/2}\log^{1/2} n)$ for both acyclic and cyclic cases under respective network sparsity conditions. The convergence of $\widehat{\mathbb{E}}\left[g_1^3(X_1)\right]$ to $\mathbb{E}\left[g_1^3(X_1)\right]$ can be established using (8.43). Recall the definitions of $\widehat{a}_i$ and a_i from (8.26) and (8.27),

$$
\widehat{\mathbb{E}}[g_1^3(X_1)] = \frac{1}{n}\sum_{i=1}^n(\widehat{a}_i - \widehat{U}_n)^3 \quad \text{and} \quad \mathbb{E}[g_1^3(X_1)] = \mathbb{E}\left[\left(\mathbb{E}[h(X_1,\cdots,X_r)|X_1] - \mu_n\right)^3\right].
$$

Observe that

$$
\left|\widehat{\mathbb{E}}[g_1^3(X_1)] - \mathbb{E}[g_1^3(X_1)]\right| \leqslant \Big|\sum_{i=1}^n(\widehat{a}_i - \widehat{U}_n)^3 - \sum_{i=1}^n(a_i-\mu_n)^3\Big|/n
$$

$$
+ \Big|\sum_{i=1}^n(a_i-\mu_n)^3/n - \mathbb{E}\big(\mathbb{E}[h(X_1,\cdots,X_r)|X_1]-\mu_n\big)^3\Big|
$$

$$
= \Big|\sum_{i=1}^n(a_i-\mu_n)^3/n - \mathbb{E}\big(\mathbb{E}[h(X_1,\cdots,X_r)|X_1]-\mu_n\big)^3\Big| + \widetilde{O}_p\Big(\rho_n^{3s-1/2}\cdot n^{-1/2}\log^{1/2} n\Big) \tag{8.78}
$$

where the last inequality is due to the facts $a_i \asymp \mu_n \asymp \rho_n^s$, $|\widehat{a}_i - a_i| = \widetilde{O}_p(\rho_n^{s-1/2}\cdot n^{-1/2}\log^{1/2} n)$ and $|\widehat{U}_n - \mu_n| = \widetilde{O}_p(\rho_n^{s-1/2}\cdot n^{-1/2}\log^{1/2} n)$ due to the proof of Lemma 3.1 (a), (b) and (c). Moreover, we have

$$
\Big|\sum_{i=1}^n(a_i-\mu_n)^3/n - \mathbb{E}\big(\mathbb{E}[h(X_1,\cdots,X_r)|X_1]-\mu_n\big)^3\Big|
$$

$$
\leqslant \Big|\sum_{i=1}^n a_i^3/n - \mathbb{E}\big(\mathbb{E}[h(X_1,\cdots,X_r)|X_i]\big)^3\Big|
$$

$$
+ \rho_n^s\cdot O\Big(\Big|\sum_{i=1}^n a_i^2/n - \mathbb{E}\big(\mathbb{E}[h(X_1,\cdots,X_r)|X_i]\big)^2\Big|\Big)
$$

$$
+ \rho_n^{2s}\cdot O\Big(\Big|\sum_{i=1}^n a_i/n - \mathbb{E}\big(\mathbb{E}[h(X_1,\cdots,X_r)|X_i]\big)\Big|\Big) \tag{8.79}
$$

Recall the definition of a_i and notice that it is a U-statistic of order $r-1$ conditioned on X_i. By the standard concentration inequality of U-statistic [97], we have

$$
\big|a_i - \mathbb{E}[h(X_1,\cdots,X_r)|X_i]\big| = \widetilde{O}_p\big(\rho_n^s\cdot n^{-1/2}\log^{1/2} n\big).
$$

By decomposing $a_i = \big(a_i - \mathbb{E}[h(X_1,\cdots,X_r)|X_i]\big) + \mathbb{E}[h(X_1,\cdots,X_r)|X_i]$, we have

$$
\rho_n^{2s}\cdot O\Big(\Big|\sum_{i=1}^n a_i/n - \mathbb{E}\big(\mathbb{E}[h(X_1,\cdots,X_r)|X_i]\big)\Big|\Big) = \widetilde{O}_p(\rho_n^{3s}\cdot n^{-1/2}\log^{1/2} n)
$$

where we used the facts $\{\mathbb{E}[h(X_1,\cdots,X_r)|X_i]\}_{i=1}^n$ are i.i.d. random variables so that

$$\left|n^{-1}\sum_{i=1}^n \mathbb{E}[h(X_1,\cdots,X_r)|X_i]-\mathbb{E}[h(X_1,\cdots,X_r)]\right|=\widetilde{O}_p(\rho_n^{3s}n^{-1/2}\log^{1/2}n).$$

By a similar strategy, we can prove that the bound $\widetilde{O}_p(\rho_n^{3s}\cdot n^{-1/2}\log^{1/2}n)$ also holds for the other two terms in RHS of (8.79). Together with (8.78), we conclude that

$$\left|\widehat{\mathbb{E}}g_1^3(X_1)-\mathbb{E}g_1^3(X_1)\right|=\widetilde{O}_p\big(\rho_n^{3s-0.5}\cdot n^{-1/2}\log^{1/2}n\big). \tag{8.80}$$

The proof of the convergence of $\widehat{\mathbb{E}}\left[g_1(X_1)g_1(X_2)g_2(X_1,X_2)\right]$, however, needs separate care. Recall that

$$\widehat{g}_1(X_i):=\frac{1}{\binom{n-1}{r-1}}\sum_{\substack{1\leqslant i_1<\cdots<i_{r-1}\leqslant n\\ i_1,\ldots,i_{r-1}\neq i}}h(A_{i,i_1,\ldots,i_{r-1}})-\widehat{U}_n=\widehat{a}_i-\widehat{U}_n$$

$$\widehat{g}_2(X_i,X_j):=\frac{1}{\binom{n-2}{r-2}}\sum_{\substack{1\leqslant i_1<\cdots<i_{r-2}\leqslant n\\ i_1,\ldots,i_{r-2}\neq i,j}}h(A_{i,j,i_1,\ldots,i_{r-2}})-\widehat{U}_n-\widehat{g}_1(X_i)-\widehat{g}_1(X_j)$$

Unlike that $\widehat{g}_1(X_i)$ converges to the corresponding $g_1(X_i)$, the randomness in $h(A_{i,j,i_1,\ldots,i_{r-2}})$ introduced by the edge A_{ij} is not suppressed by an average over $\{i_1,\ldots,i_{r-2}\}:i_1,\ldots,i_{r-2}\neq i,j$. Therefore, the convergence of $\widehat{\mathbb{E}}\left[g_1(X_1)g_1(X_2)g_2(X_1,X_2)\right]$ has to be discussed as a whole. We first show that given W, $\widehat{\mathbb{E}}\left[g_1(X_1)g_1(X_2)g_2(X_1,X_2)\right]$ converges to its "population-sample" version replacing A by W in its definition, then show the convergence of that version to the eventual expectation form. Observe that

$$\begin{aligned}
&\frac{1}{\binom{n}{2}}\sum_{1\leqslant i<j\leqslant n}\widehat{g}_1(X_i)\widehat{g}_1(X_j)\widehat{g}_2(X_i,X_j)-\mathbb{E}g_1(X_1)g_1(X_2)g_2(X_1,X_2)\\
=&\frac{1}{\binom{n}{2}}\sum_{1\leqslant i<j\leqslant n}\left[\widehat{g}_1(X_i)\widehat{g}_1(X_j)\widehat{g}_2(X_i,X_j)-g_1(X_i)g_1(X_j)g_2(X_i,X_j)\right]\\
&+\frac{1}{\binom{n}{2}}\sum_{1\leqslant i<j\leqslant n}g_1(X_i)g_1(X_j)g_2(X_i,X_j)-\mathbb{E}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)].
\end{aligned}$$

It is easy to bound the second term. By the definition of $g_1(X_i),g_2(X_i,X_j)$, we notice that clearly $\binom{n}{2}^{-1}\sum_{1\leqslant i<j\leqslant n}g_1(X_i)g_1(X_j)g_2(X_i,X_j)$ is a degree-two U-statistic. By the standard concentration inequality of U-statistic [97],

$$\begin{aligned}
&\left|\frac{1}{\binom{n}{2}}\sum_{1\leqslant i<j\leqslant n}g_1(X_i)g_1(X_j)g_2(X_i,X_j)-\mathbb{E}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)]\right|\\
&\qquad=\widetilde{O}_p\big(\rho_n^{3s}n^{-1/2}\log^{1/2}n\big)
\end{aligned}$$

where we used the fact $g_1(X_i)g_1(X_j)g_2(X_i,X_j)=O\big(\rho_n^{3s}\big)$ a.s. Therefore, it suffices to upper bound

$$\mathfrak{K}_1:=\frac{1}{\binom{n}{2}}\sum_{1\leqslant i<j\leqslant n}\big[\widehat{g}_1(X_i)\widehat{g}_1(X_j)\widehat{g}_2(X_i,X_j)-g_1(X_i)g_1(X_j)g_2(X_i,X_j)\big]. \tag{8.81}$$

The convergence of $\widehat{g}_1(X_i)$ to $g_1(X_i)$ is straightforward. Indeed,

$$\widehat{g}_1(X_i)-g_1(X_i)=\widehat{a}_i-\mathbb{E}[h(X_1,\cdots,X_r)|X_i]+(\mu_n-\widehat{U}_n).$$

Recall from Lemma 3.1(a), (b) and (c), $|\widehat{U}_n - \mu_n| = \widetilde{O}_p(\rho_n^{s-1/2} \cdot n^{-1/2} \log^{1/2} n)$. We then prove the first term on RHS of above equation. Clearly,

$$\begin{aligned}|\widehat{a}_i - \mathbb{E}[h(X_1, \cdots, X_r)|X_i]| &\leqslant |\widehat{a}_i - a_i| + |a_i - \mathbb{E}[h(X_1, \cdots, X_r)|X_i]| \\ &= \widetilde{O}_p(\rho_n^{s-1/2} \cdot n^{-1/2} \log^{1/2} n)\end{aligned}$$

where the last inequality is due to the bounds of $|\widehat{a}_i - a_i|$ and $|a_i - \mathbb{E}[h(X_1, \cdots, X_r)|X_i]|$ as shown above. Therefore, conditioned on X_i, we have $|\widehat{g}_1(X_i) - g_1(X_i)| = \widetilde{O}_p(\rho_n^{s-1/2} \cdot n^{-1/2} \log^{1/2} n)$. Now, we re-express $\mathfrak{K}_1$ from (8.81) as

$$\begin{aligned}\mathfrak{K}_1 =& \frac{1}{\binom{n}{2}} \sum_{1 \leqslant i < j \leqslant n} \widehat{g}_1(X_i)\widehat{g}_1(X_j)\big[\widehat{g}_2(X_i, X_j) - g_2(X_i, X_j)\big] \\ &+ \frac{1}{\binom{n}{2}} \sum_{1 \leqslant i < j \leqslant n} \big[\widehat{g}_1(X_i)\widehat{g}_1(X_j)g_2(X_i, X_j) - g_1(X_i)g_1(X_j)g_2(X_i, X_j)\big] \\ =& \frac{1}{\binom{n}{2}} \sum_{1 \leqslant i < j \leqslant n} \widehat{g}_1(X_i)\widehat{g}_1(X_j)\big[\widehat{g}_2(X_i, X_j) - g_2(X_i, X_j)\big] + \widetilde{O}_p(\rho_n^{3s-1/2} \cdot n^{-1/2} \log^{1/2} n),\end{aligned}$$

where we used the fact $|g_1(X_i)| = O(\rho_n^s)$, a.s. It suffices to bound the first term on RHS. Define

$$\widehat{a}_{ij} := \frac{1}{\binom{n-2}{r-2}} \sum_{\substack{1 \leqslant i_1 < i_2 < \cdots < i_{r-2} \leqslant n \\ i_1, \cdots, i_{r-2} \neq i, j}} h(A_{i,j,i_1,i_2,\cdots,i_{r-2}}) \tag{8.82}$$

$$a_{ij} := \frac{1}{\binom{n-2}{r-2}} \sum_{\substack{1 \leqslant i_1 < i_2 < \cdots < i_{r-2} \leqslant n \\ i_1, \cdots, i_{r-2} \neq i, j}} h(W_{i,j,i_1,i_2,\cdots,i_{r-2}}).$$

Then we can re-express the $\widehat{g}_2(X_i, X_j) - g_2(X_i, X_j)$ factor as follows

$$\begin{aligned}\widehat{g}_2(X_i, X_j) - g_2(X_i, X_j) &= (\widehat{a}_{ij} - a_{ij}) + \big(a_{ij} - \mathbb{E}[h(X_1, \cdots, X_r)|X_i, X_j]\big) \\ &\quad - (\widehat{U}_n - \mu_n) - \big(\widehat{g}_1(X_i) - g_1(X_i)\big) - \big(\widehat{g}_1(X_j) - g_1(X_j)\big).\end{aligned}$$

Similarly to our earlier derivations, using the concentration of U-statistics, we have $\big(a_{ij} - \mathbb{E}[h(X_1, \cdots, X_r)|X_i, X_j]\big) = \widetilde{O}_p\big(\rho_n^s n^{-1/2} \log^{1/2} n\big)$. Since $\widehat{U}_n - \mu_n = \widetilde{O}_p(\rho_n^{s-1/2} n^{-1/2} \log^{1/2} n)$ and $\widehat{g}_1(X_i) - g_1(X_i) = \widetilde{O}_p(\rho_n^{s-1/2} \cdot n^{-1/2} \log^{1/2} n)$, we have

$$\begin{aligned}&\frac{1}{\binom{n}{2}} \sum_{1 \leqslant i < j \leqslant n} \widehat{g}_1(X_i)\widehat{g}_1(X_j)\big[\widehat{g}_2(X_i, X_j) - g_2(X_i, X_j)\big] \\ &\qquad = \frac{1}{\binom{n}{2}} \sum_{1 \leqslant i < j \leqslant n} \widehat{g}_1(X_i)\widehat{g}_1(X_j)(\widehat{a}_{ij} - a_{ij}) + \widetilde{O}_p(\rho_n^{3s-1/2} \cdot n^{-1/2} \log^{1/2} n).\end{aligned}$$

Therefore, we have

$$\begin{aligned}&\frac{1}{\binom{n}{2}} \sum_{1 \leqslant i < j \leqslant n} \widehat{g}_1(X_i)\widehat{g}_1(X_j)\widehat{g}_2(X_i, X_j) - \mathbb{E}g_1(X_1)g_1(X_2)g_2(X_1, X_2) \\ &\qquad = \frac{1}{\binom{n}{2}} \sum_{1 \leqslant i < j \leqslant n} \widehat{g}_1(X_i)\widehat{g}_1(X_j)(\widehat{a}_{ij} - a_{ij}) + \widetilde{O}_p(\rho_n^{3s-1/2} \cdot n^{-1/2} \log^{1/2} n).\end{aligned}$$

Recall the definitions of $\widehat{a}_i$ and a_i from (8.26) and (8.27). We write

$$\begin{aligned}
&\frac{1}{\binom{n}{2}}\sum_{1\leqslant i<j\leqslant n}\widehat{g}_1(X_i)\widehat{g}_1(X_j)(\widehat{a}_{ij}-a_{ij}) = \frac{1}{\binom{n}{2}}\sum_{1\leqslant i<j\leqslant n}\widehat{a}_i\widehat{a}_j(\widehat{a}_{ij}-a_{ij})\\
&\qquad\qquad\qquad\qquad -\frac{2}{n}\sum_{1\leqslant i\leqslant n}\widehat{U}_n\widehat{a}_i(\widehat{a}_i-a_i)+\widehat{U}_n^2(\widehat{U}_n-U_n)\\
&\qquad\qquad\qquad\qquad =\frac{1}{\binom{n}{2}}\sum_{1\leqslant i<j\leqslant n}\widehat{a}_i\widehat{a}_j(\widehat{a}_{ij}-a_{ij})+\widetilde{O}_p(\rho_n^{3s-1/2}\cdot n^{-1/2}\log^{1/2}n)
\end{aligned}$$

where the last equation is due to $a_i \asymp U_n \asymp \rho_n^s$ a.s., $|\widehat{a}_i - a_i| = \widetilde{O}_p(\rho_n^{s-1/2}\cdot n^{-1/2}\log^{1/2} n)$, $|\widehat{U}_n - U_n| = \widetilde{O}_p(\rho_n^{s-1/2}\cdot n^{-1}\log^{1/2} n)$ due to Lemma 3.1 (b). Therefore,

$$\begin{aligned}
&\frac{1}{\binom{n}{2}}\sum_{1\leqslant i<j\leqslant n}\widehat{g}_1(X_i)\widehat{g}_1(X_j)\widehat{g}_2(X_i,X_j)-\mathbb{E}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)]\\
&\qquad =\frac{1}{\binom{n}{2}}\sum_{1\leqslant i<j\leqslant n}\widehat{a}_i\widehat{a}_j(\widehat{a}_{ij}-a_{ij})+\widetilde{O}_p(\rho_n^{3s-1/2}\cdot n^{-1/2}\log^{1/2}n).
\end{aligned}\tag{8.83}$$

It remains to bound the first term on RHS. We rewrite it as

$$\begin{aligned}
\frac{1}{\binom{n}{2}}\sum_{1\leqslant i<j\leqslant n}\widehat{a}_i\widehat{a}_j(\widehat{a}_{ij}-a_{ij}) &= \frac{1}{n(n-1)}\sum_{1\leqslant i\neq j\leqslant n}\widehat{a}_i\widehat{a}_j(\widehat{a}_{ij}-a_{ij})\\
&=\frac{1}{n}\sum_{i=1}^n\widehat{a}_i\cdot\Big(\frac{1}{n-1}\sum_{j\neq i}\widehat{a}_j(\widehat{a}_{ij}-a_{ij})\Big).
\end{aligned}\tag{8.84}$$

We then establish the upper bound for $\sum_{j\neq i}\widehat{a}_j(\widehat{a}_{ij}-a_{ij})/(n-1)$ for each fixed i. We have

$$\begin{aligned}
&\frac{1}{n-1}\sum_{\substack{1\leqslant j\leqslant n\\ j\neq i}}(\widehat{a}_j-a_j)(\widehat{a}_{ij}-a_{ij})\\
&\qquad =\frac{1}{(n-1)^2}\sum_{\substack{1\leqslant j\leqslant n\\ j\neq i}}\sum_{\substack{1\leqslant i'\leqslant n\\ i'\neq j}}(\widehat{a}_{i'j}-a_{i'j})(\widehat{a}_{ij}-a_{ij})\\
&\qquad =\frac{1}{(n-1)^2}\Bigg\{\sum_{\substack{1\leqslant j\leqslant n\\ j\neq i}}(\widehat{a}_{ij}-a_{ij})^2+\sum_{\substack{1\leqslant\{i',j\}\leqslant n\\ i'\neq i\\ j\neq i,i'}}(\widehat{a}_{i'j}-a_{i'j})(\widehat{a}_{ij}-a_{ij})\Bigg\}
\end{aligned}\tag{8.85}$$

Similar to the derivation of (8.43) by expanding $\sum_{\substack{1\leqslant i_1<\cdots<i_r\leqslant n\\ \{i,j\}\subset\{i_1,\cdots,i_r\}}} h(A_{i_1,\cdots,i_r})$, we have

$$\begin{aligned}
\widehat{a}_{ij} &= \mathring{\Theta}_{ij}\eta_{ij}+\frac{1}{n-2}\sum_{\substack{1\leqslant k\leqslant n\\ k\neq i,j}}\Big(\mathring{\Theta}_{i,k;i,j}\eta_{ik}+\mathring{\Theta}_{j,k;i,j}\eta_{jk}\Big)\\
&\quad + a_{ij}+\widetilde{O}_p(\rho_n^{s-1}\cdot n^{-1}\log n)
\end{aligned}\tag{8.86}$$

where

$$\max\big\{|\mathring{\Theta}_{ij}|,|\mathring{\Theta}_{i,k;i,j}|,|\mathring{\Theta}_{j,k;i,j}|\big\}\preceq\rho_n^{s-1}, a.s.$$

We note that , similarly as the derivation of (8.43), the bound (8.86) holds under the sparsity condition $\rho_n=\omega(n^{-1})$ for acyclic R and $\rho_n=\omega(n^{-2/r})$ for cyclic R.

Now we discuss the two terms on the RHS of (8.85). For term 1 on the RHS of (8.85), we have

$$
\begin{aligned}
&\frac{1}{(n-1)^2}\sum_{\substack{1\leqslant j\leqslant n\\ j\neq i}}(\widehat{a}_{ij}-a_{ij})^2\\
&\qquad=\frac{1}{(n-1)^2}\sum_{\substack{1\leqslant j\leqslant n\\ j\neq i}}\left\{\mathring{\Theta}_{i,j}\eta_{ij}+\frac{1}{n-2}\sum_{\substack{1\leqslant k\leqslant n\\ k\neq i,j}}\left(\mathring{\Theta}_{i,k;i,j}\eta_{ik}+\mathring{\Theta}_{j,k;i,j}\eta_{jk}\right)\right\}^2\\
&\qquad\quad+\widetilde{O}_p(\rho_n^{2s-1}\cdot n^{-2}\log n)\\
&\qquad\asymp n^{-2}\left\{\sum_{\substack{1\leqslant j\leqslant n\\ j\neq i}}\mathring{\Theta}_{i,j}^2\eta_{ij}^2+\frac{2}{n-2}\sum_{\substack{1\leqslant \{j,k\}\leqslant n\\ j\neq i\\ k\neq i,j}}\mathring{\Theta}_{i,j}\eta_{ij}\left(\mathring{\Theta}_{i,k;i,j}\eta_{ik}+\mathring{\Theta}_{j,k;i,j}\eta_{jk}\right)\right.\\
&\qquad\quad\left.+\frac{1}{(n-2)^2}\sum_{\substack{1\leqslant j\leqslant n\\ j\neq i}}\left(\sum_{\substack{1\leqslant k\leqslant n\\ k\neq i,j}}\left(\mathring{\Theta}_{i,k;i,j}\eta_{ik}+\mathring{\Theta}_{j,k;i,k}\eta_{jk}\right)\right)^2\right\}+\widetilde{O}_p(\rho_n^{2s-1}\cdot n^{-2}\log n)
\end{aligned}
\tag{8.87}
$$

Now we bound each term on the RHS of (8.87). Inspecting the expectation of term 1 on the RHS of (8.87) and using Bernstein inequality, we know it is $\widetilde{O}_p(\rho_n^{2s-1}\cdot n^{-1}+\rho_n^{2s-3/2}\cdot n^{-3/2}\log^{1/2}n)$. Term 2 on the RHS of (8.87) is mean zero so we can focus on the concentration. Its $\eta_{ij}\eta_{ik}$ part can be bounded by inspecting the concentration averaging over j and over k, respectively, and see that this part is bounded as $\widetilde{O}_p(\rho_n^{2s-2}\cdot n^{-1}(\rho_n n^{-1/2}\log^{1/2}n)^2)$, and this upper bound is dominated by the bound of term 1, thus it is ignorable. Using Theorem 8.1, the $\eta_{ij}\eta_{jk}$ part of term 2 can be bounded as follows

$$
n^{-3}\sum_{\substack{1\leqslant \{j,k\}\leqslant n\\ j\neq i\\ k\neq i,j}}\mathring{\Theta}_{i,j}\mathring{\Theta}_{j,k;i,j}\eta_{ij}\eta_{jk}=\widetilde{O}_p\left(n^{-3}\cdot\rho_n^{2s-2}\cdot\max\left\{\underbrace{\sqrt{\rho_n^2\cdot n^2\log n}}_{\text{“Variance”}},\underbrace{\rho_n\cdot n\log n}_{\text{“}\Xi_1\text{”}}\right\}\right)
$$

and is thus ignorable. Now noticing that each η is bounded by 1, using Bernstein's inequality, term 3 on the RHS of (8.87) is $\widetilde{O}_p(n^{-4}\cdot\rho_n^{2s-1}\cdot n^2)=\widetilde{O}_p(\rho_n^{2s-1}\cdot n^{-2})$ and thus ignorable. Therefore, term 1 on the RHS of (8.85) is $\widetilde{O}_p(\rho_n^{2s-1}\cdot n^{-1}\log n+\rho_n^{2s-3/2}\cdot n^{-3/2}\log^{1/2}n)$.

Now we bound term 2 on the RHS of (8.85). By a similar treatment, we have

$$
\begin{aligned}
&\frac{1}{(n-1)^2}\sum_{\substack{1\leqslant \{i',j\}\leqslant n\\ i'\neq i\\ j\neq i,i'}}(\widehat{a}_{i'j}-a_{i'j})(\widehat{a}_{ij}-a_{ij})\\
&\qquad=\frac{1}{(n-1)^2}\sum_{\substack{1\leqslant \{i',j\}\leqslant n\\ i'\neq i\\ j\neq i,i'}}\left\{\mathring{\Theta}_{i',j}\eta_{i'j}+\frac{1}{n-2}\sum_{\substack{1\leqslant k_1\leqslant n\\ k_1\neq i',j}}\left(\mathring{\Theta}_{i',k_1;i',j}\eta_{i'k}+\mathring{\Theta}_{j,k_1;i',j}\eta_{jk_1}\right)\right\}
\end{aligned}
$$

$$\cdot\left\{\mathring{\Theta}_{ij}\eta_{ij}+\frac{1}{n-2}\sum_{\substack{1\leqslant k_2\leqslant n\\k_2\neq i,j}}\left(\mathring{\Theta}_{i,k_2;i,j}\eta_{ik_2}+\mathring{\Theta}_{j,k_2;i,j}\eta_{jk_2}\right)\right\}+\widetilde{O}_p(\rho_n^{2s-1}\cdot n^{-1}\log n)$$

$$\asymp n^{-2}\Bigg[\sum_{\substack{1\leqslant\{i',j\}\leqslant n\\i'\neq i\\j\neq i,i'}}\mathring{\Theta}_{i',j}\mathring{\Theta}_{i,j}\eta_{i'j}\eta_{ij}+\frac{1}{n-2}\sum_{\substack{1\leqslant\{i',j,k_2\}\leqslant n\\i'\neq i\\j\neq i,i'\\k_2\neq i,j}}\left(\mathring{\Theta}_{i',j}\mathring{\Theta}_{i,k_2;i,j}\eta_{i'j}\eta_{ik_2}+\mathring{\Theta}_{i',j}\mathring{\Theta}_{j,k_2;i,j}\eta_{i'j}\eta_{jk_2}\right)$$

$$+\frac{1}{n-2}\sum_{\substack{1\leqslant\{i',j,k_1\}\leqslant n\\i'\neq i\\j\neq i,i'\\k_1\neq i,j}}\left(\mathring{\Theta}_{i,j}\mathring{\Theta}_{i',k_1;i',j}\eta_{ij}\eta_{i'k_1}+\mathring{\Theta}_{i,j}\mathring{\Theta}_{j,k_1;i',j}\eta_{ij}\eta_{jk_1}\right)$$

$$+\widetilde{O}_p((\rho_n^{s-1/2}\cdot n^{-1/2}\log^{1/2}n)^2)+\widetilde{O}_p(\rho_n^{2s-1}\cdot n^{-1}\log n)\Bigg]\tag{8.88}$$

Now we bound the RHS of (8.88). Again, by Theorem 8.1, the first term is bounded by

$$n^{-2}\sum_{\substack{1\leqslant\{i',j\}\leqslant n\\i'\neq i\\j\neq i,i'}}\mathring{\Theta}_{i',j}\eta_{i'j}=\widetilde{O}_p(\rho_n^{2s-1}\cdot n^{-1}\log n).\tag{8.89}$$

Terms 2 and 3 on the RHS of (8.88) can be bounded exactly similarly. Here we only present the bounding of term 2. We have

$$\frac{1}{(n-2)^3}\sum_{\substack{1\leqslant\{i',j,k_2\}\leqslant n\\i'\neq i\\j\neq i,i'\\k_2\neq i,j}}\mathring{\Theta}_{i',j}\mathring{\Theta}_{i,k_2;i,j}\eta_{i'j}\eta_{ik_2}=\frac{1}{(n-2)^3}\sum_{1\leqslant i'\leqslant n}\left(\sum_{j\neq i,i'}\mathring{\Theta}_{i',j}\eta_{i'j}\sum_{k_2\neq i,j}\mathring{\Theta}_{i,k_2;i,j}\eta_{ik_2}\right)$$

$$=n^{-3}\rho_n^{2s-2}\widetilde{O}_p((\rho_n^{1/2}n^{1/2}\log^{1/2})^2)=\widetilde{O}_p(\rho_n^{2s-1}\cdot n^{-2}\log n)\tag{8.90}$$

and using Theorem 8.1, we have

$$\frac{1}{(n-2)^3}\sum_{\substack{1\leqslant\{i',j,k_2\}\leqslant n\\i'\neq i\\j\neq i,i'\\k_2\neq i,j}}\mathring{\Theta}_{i',j}\mathring{\Theta}_{j,k_2;i,j}\eta_{i'j}\eta_{jk_2}$$

$$=n^{-3}\rho_n^{2s-2}\cdot\widetilde{O}_p(\max\{\rho_n\cdot n^{3/2}\log^{1/2}n,\rho_n\cdot n\log n\})=\widetilde{O}_p(\rho_n^{2s-1}\cdot n^{-3/2}\log n)\tag{8.91}$$

Collecting all results, we see that term 2 on the RHS of (8.85) is $\widetilde{O}_p(\rho_n^{2s-1}\cdot n^{-1}\log n+\rho_n^{2s-3/2}\cdot n^{-3/2}\log n)$. We thus conclude that

$$\left|\frac{1}{n-1}\sum_{\substack{1\leqslant j\leqslant n\\j\neq i}}(\widehat{a}_{ij}-a_{ij})(\widehat{a}_j-a_j)\right|=\widetilde{O}_p(\rho_n^{2s-1}n^{-1}\log n)$$

under the given sparsity condition $\rho_n = \omega(n^{-1/2})$, which holds for both acyclic and cyclic R.

Now we return to the main proof and continue (8.84). We have

$$
\begin{aligned}
&\frac{1}{\binom{n}{2}} \sum_{1\leqslant i<j\leqslant n} \widehat{a}_i \widehat{a}_j (\widehat{a}_{ij} - a_{ij}) \\
&\qquad = \frac{1}{\binom{n}{2}} \sum_{1\leqslant i<j\leqslant n} \widehat{a}_i a_j (\widehat{a}_{ij} - a_{ij}) + \frac{1}{\binom{n}{2}} \sum_{1\leqslant i<j\leqslant n} \widehat{a}_i (\widehat{a}_j - a_j)(\widehat{a}_{ij} - a_{ij}) \\
&\qquad = \frac{1}{\binom{n}{2}} \sum_{1\leqslant i<j\leqslant n} a_i a_j (\widehat{a}_{ij} - a_{ij}) + \frac{1}{\binom{n}{2}} \sum_{1\leqslant i<j\leqslant n} (\widehat{a}_i - a_i) a_j (\widehat{a}_{ij} - a_{ij}) \\
&\qquad\quad + \widetilde{O}_p(\rho_n^{3s-1} \cdot n^{-1} \log n + \rho_n^{3s-3/2} \cdot n^{-3/2} \log n) \\
&\qquad = \widetilde{O}_p(\rho_n^{3s-1} \cdot n^{-1} \log n) + \widetilde{O}_p(\rho_n^{3s-1} \cdot n^{-1} \log n + \rho_n^{3s-3/2} \cdot n^{-3/2} \log n) \\
&\qquad = \widetilde{O}_p(\rho_n^{3s-1} \cdot n^{-1} \log n + \rho_n^{3s-3/2} \cdot n^{-3/2} \log n) \qquad (8.92)
\end{aligned}
$$

where the second to last line is due to

$$
\begin{aligned}
&\frac{1}{\binom{n}{2}} \sum_{1\leqslant i<j\leqslant n} a_i a_j (\widehat{a}_{ij} - a_{ij}) \\
&\qquad = \frac{1}{\binom{n}{2}} \sum_{1\leqslant i<j\leqslant n} a_i a_j \left\{ \mathring{\Theta}_{i,j} \eta_{ij} + \frac{1}{n-2} \sum_{\substack{1\leqslant k\leqslant n \\ k\neq i,j}} \left(\mathring{\Theta}_{i,k;i,j} \eta_{ik} + \mathring{\Theta}_{j,k;i,j} \eta_{jk} \right) \right\} \\
&\qquad\quad + \widetilde{O}_p(\rho_n^{3s-1} \cdot n^{-1} \log n) \\
&\qquad = \frac{1}{\binom{n}{2}} \sum_{1\leqslant i<j\leqslant n} a_i a_j \mathring{\Theta}_{i,j} \eta_{ij} + \frac{1}{\binom{n}{2}} \sum_{\substack{1\leqslant \{i,j,k\}\leqslant n \\ i\neq j; j\neq k; k\neq i}} a_i a_j \frac{\mathring{\Theta}_{i,k;i,j} \eta_{ik}}{n-2} + \widetilde{O}_p(\rho_n^{3s-1} \cdot n^{-1} \log n) \\
&(\text{Bernstein}) = \widetilde{O}_p(\rho_n^{3s-1} \cdot n^{-1} \log n) + \frac{1}{\binom{n}{2}} \sum_{\substack{1\leqslant \{i,k\}\leqslant n \\ i\neq k}} a_i \left(\sum_{\substack{1\leqslant j\leqslant n \\ j\neq i,k}} a_j \frac{\mathring{\Theta}_{i,k;i,j}}{n-2} \right) \eta_{ik} \\
&(8.93) \\
&\qquad = \widetilde{O}_p(\rho_n^{3s-1} \cdot n^{-1} \log n)
\end{aligned}
$$

Now we may conclude that

$$
\begin{aligned}
&\left| \frac{1}{\binom{n}{2}} \sum_{1\leqslant i<j\leqslant n} \widehat{g}_1(X_i) \widehat{g}_1(X_j) \widehat{g}_2(X_i, X_j) - \mathbb{E}[g_1(X_1) g_1(X_2) g_2(X_1, X_2)] \right| \\
&\qquad = \widetilde{O}_p(\rho_n^{3s-1/2} \cdot n^{-1/2} \log^{1/2} n) \qquad (8.94)
\end{aligned}
$$

This completes the proof of Theorem 3.2.

8.5. *Proof of Theorem 3.3.* We will inherit the notation of $\widehat{a}_i$ from (8.27) in the proof of Lemma 3.1. It suffices to show (3.15), which would then imply the closeness between $F_{\widehat{T}_n}$

and $F_{\widehat{T}_{n;\text{bootstrap}}}$ by repeating our arguments for proving (8.54) and (8.55) using Lemma 8.2. Observe that

$$\binom{n}{r}\cdot \widehat{U}_n = \sum_{1\leqslant i_1<\ldots<i_r\leqslant n} h(A_{i_1,\ldots,i_r})$$
$$\text{(For any } i) = \sum_{\substack{1\leqslant i_1<\ldots<i_{r-1}\leqslant n\\ i_1,\ldots,i_{r-1}\neq i}} h(A_{i,i_1,\ldots,i_{r-1}}) + \sum_{\substack{1\leqslant i_1<\ldots<i_r\leqslant n\\ i_1,\ldots,i_r\neq i}} h(A_{i_1,\ldots,i_r})$$
$$= \binom{n-1}{r-1}\cdot \widehat{a}_i + \binom{n-1}{r}\cdot \widehat{U}_n^{(-i)}$$

Simplifying both sides, we have

$$\widehat{U}_n^{(-i)} - \widehat{U}_n = -\frac{r}{n-r}\left(\widehat{a}_i - \widehat{U}_n\right) \tag{8.95}$$

Therefore,

$$n\left(\widehat{S}_n^2 - \widehat{S}^2_{n;\text{jackknife}}\right)$$
$$=\frac{r^2}{n}\sum_{i=1}^n(\widehat{a}_i-\widehat{U}_n)^2 - (n-1)\sum_{i=1}^n\left(\widehat{U}_n^{(-i)}-\widehat{U}_n\right)^2$$
$$=\frac{1}{n}\sum_{i=1}^n\left[r^2\left(\widehat{a}_i-\widehat{U}_n\right)^2 - n(n-1)\cdot\frac{r^2}{(n-r)^2}\left(\widehat{a}_i-\widehat{U}_n\right)^2\right]$$
$$=\frac{1}{n}\sum_{i=1}^n r^2\left\{1-\frac{n(n-1)}{(n-r)^2}\right\}\left(\widehat{a}_i-\widehat{U}_n\right)^2 = O(\widehat{S}_n^2) \tag{8.96}$$

where in the last line, recall that $\widehat{S}_n^2 := r^2\sum_{i=1}^n(\widehat{a}_i-\widehat{U}_n)^2/n^2$. Therefore,

$$\widehat{S}_n^2 - \widehat{S}^2_{n;\text{jackknife}} = O(\widehat{S}_n^2/n) \implies |\widehat{S}_n - \widehat{S}_{n;\text{jackknife}}| = O(\widehat{S}_n/n).$$

This proves (3.15) and thus completes the proof of Theorem 3.3.

PROOF OF THEOREM 3.4. It suffices to prove the Berry-Esseen bound for the normal approximation. By definition, $\|G_n(u)-\Phi(u)\|_\infty = O(n^{-1/2})$ and by the proof of Theorem 3.2, we know that $\|\widehat{G}_n(u)-G_n(u)\|_\infty = \widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)\cdot n^{-1/2} = \widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)\bigwedge o_p(1)$. The proof is partitioned into two parts, for $O\left(\mathcal{M}(\rho_n,n;R)\right)$ and $o(1)$ bounds, respectively.

Part I: proof of the $O\left(\mathcal{M}(\rho_n,n;R)\right)$ bound when $\rho_n n \succ \log^{1/2} n$ (acyclic) or $\rho_n^{r/2} n \succ \log^{1/2} n$ (cyclic).

We begin by recalling the decomposition of $\widehat{T}_n$ and inspect whether each component depends on ρ_n or not. Using Lemma 3.1 for the sparse regime, we have

$$\widehat{T}_n = \left\{\underbrace{U_n^\# + \Delta_n}_{\text{No }\rho_n} + \underbrace{\widehat{\Delta}_n}_{\text{Depend on }\rho_n} + \underbrace{\widetilde{O}_p(n^{-1}\log^{3/2} n)}_{\text{No }\rho_n}\right\}\cdot\left(1+\underbrace{\delta_n}_{\text{No }\rho_n}+\underbrace{\widehat{\delta}_n}_{\text{Depend on }\rho_n}\right)$$
$$=\left\{U_n^\# + \widetilde{O}_p(n^{-1/2}\log^{1/2} n) + \widehat{\Delta}_n + \widetilde{O}_p(n^{-1}\log^{3/2} n)\right\}$$
$$\cdot\left\{1+\widetilde{O}_p(n^{-1/2}\log^{1/2} n) + \widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)\right\}$$

$$=U_n^{\#}+\widehat{\Delta}_n\big(1+\widetilde{O}_p(\mathcal{M}(\rho_n,n;R))\big)+\widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right) \tag{8.97}$$

To see that last equality of (8.97), it is not difficult to prove that $U_n^{\#}\cdot\widehat{\delta}_n$ is also $\widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)$ using the method for proving that $\widehat{\delta}_n=\widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)$, but due to its involvement we omit the proof here.

Now we use (8.97) for this proof. First, we discuss the term $\widehat{\Delta}_n\cdot\widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)$. By an ordinary Bernstein's inequality, we have

$$\mathbb{P}(|\check{\Delta}_n|>u)\leqslant 2\exp\left\{-\frac{C_1u^2n^4}{C_2n^2\cdot\rho_n\cdot\rho_n^{-2}\cdot n+C_3\rho_n^{-1}\cdot n^{1/2}\cdot u\cdot n^2}\right\}$$

$$\leqslant 2\exp\left\{-C_4(\rho_n\cdot n)\cdot u^2\right\}$$

Therefore, $\check{\Delta}_n=\widetilde{O}_p((\rho_n\cdot n)^{-1/2}\log^{1/2}n)$. Therefore, we have

$$\widehat{\Delta}_n\cdot\widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)$$

$$=(\check{\Delta}_n+\check{R}_n)\cdot\widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)$$

$$=\left\{\widetilde{O}_p((\rho_n\cdot n)^{-1/2}\log^{1/2}n)+\widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)\right\}\cdot\widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)$$

Therefore, the term $\widehat{\Delta}_n\cdot\widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)$ is ignorable compared to $\widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)$. Thus, recalling $\check{R}_n=\widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)$, we have

$$\widehat{T}_n=U_n^{\#}+\check{\Delta}_n+\widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right)$$

Now it only remains to show that

$$\|F_{U_n^{\#}+\check{\Delta}_n+\widetilde{O}_p(\mathcal{M}(\rho_n,n;R))}(u)-\Phi(u)\|_\infty=O\left(\mathcal{M}(\rho_n,n;R)\right)$$

Similar to the proof of Theorem 3.1, we are going to break this down into three steps. Recall the definition of $\widetilde{\Delta}_n$ from the proof of Theorem 3.1, we shall prove

$$\left\|F_{U_n^{\#}+\check{\Delta}_n+\widetilde{O}_p(\mathcal{M}(\rho_n,n;R))}(u)-F_{U_n^{\#}+\check{\Delta}_n}(u)\right\|_\infty=\widetilde{O}_p\left(\mathcal{M}(\rho_n,n;R)\right) \tag{8.98}$$

$$\left\|F_{U_n^{\#}+\check{\Delta}_n}(u)-F_{U_n^{\#}+\widetilde{\Delta}_n}(u)\right\|_\infty=O(\rho_n^{-1/2}\cdot n^{-1}) \tag{8.99}$$

$$\left\|F_{U_n^{\#}+\widetilde{\Delta}_n}(u)-\Phi(u)\right\|_\infty=O((\rho_n\cdot n)^{-1}\log^{1/2}n) \tag{8.100}$$

We start from proving (8.100). Notice that this part of the proof only requires that $\rho_n n\succ\log^{1/2}n$ regardless of the shape of the motif, since the asymptotic orders $U_n^{\#}\asymp 1$ and $\widetilde{\Delta}_n\asymp(\rho_n\cdot n)^{-1/2}$ do not depend on the motif. The stronger condition $\rho_n^{r/2}n\succ\log^{1/2}n$ is still necessary to deduce (8.98) from (8.99) and (8.100) using Lemma 8.2; a second reason is that the error bound $\mathcal{M}(\rho_n,n;R)$ for cyclic motifs would not diminish to zero if $\rho_n^{r/2}n\preceq\log^{1/2}n$. We are going to apply the Esseen's smoothing lemma on the interval $t\in[-\rho_n\cdot n\log^{-1/2}n,\rho_n\cdot n\log^{-1/2}n]$. The integral we shall need to bound is

$$\int_{-\rho_n\cdot n\log^{-1/2}n}^{\rho_n\cdot n\log^{-1/2}n}\left|\frac{\mathbb{E}[e^{\mathbb{i}t(U_n^{\#}+\widetilde{\Delta}_n)}]-e^{-t^2/2}}{t}\right|\mathrm{d}t \tag{8.101}$$

The following intermediate result in the proof of Lemma 8.3-(c) remains valid:

$$\mathbb{E}\left[e^{\mathbb{i}t(U_n^{\#}+\widetilde{\Delta}_n)}\right]=\mathbb{E}\left[e^{\mathbb{i}tU_n^{\#}}\cdot e^{-(\rho_n\cdot n)^{-1}\sigma_w^2t^2/2}\right]$$

$$=\mathbb{E}\left[e^{\mathbb{i}tU_n^{\#}}\cdot e^{-(\rho_n\cdot n)^{-1}t^2/2\{\mathbb{E}[\sigma_w^2]+\frac{1}{n}\sum_{i=1}^n g_{\sigma;1}(X_i)\}}\cdot(1+\widetilde{O}_p(\rho_n^{-1}n^{-2}\log n\cdot t^2))\right]$$

For $t \leqslant \rho_n \cdot n \log^{-1/2} n \ll n^{1/2}$, the remainder's contribution to the integral (8.101) is

$$\int_0^{\rho_n \cdot n \log^{-1/2} n} \rho_n^{-1} n^{-2} \log n \cdot t^2/t \mathrm{d}t = O(\rho_n) \ll n^{-1/2} \tag{8.102}$$

Therefore, for the rest of the proof in this part, we can directly ignore the remainder term's contribution according to (8.102). Now we bound the main part. Suppose $C_0 > 0$ is a very large constant. We discuss two cases

- **Case 1:** $\rho_n \cdot n \log^{-1/2} n \geqslant \{C_0 \log(\rho_n \cdot n)\}^{1/2}$. In this case, we break the integral in (8.101) into two parts:

$$\int_0^{\{C_0 \log(\rho_n \cdot n)\}^{1/2}} + \int_{\{C_0 \log(\rho_n \cdot n)\}^{1/2}}^{\rho_n \cdot n \log^{-1/2} n}$$

By (8.102), we can ignore the remainder. Similar to the intermediate step in the proof of Lemma 8.3-(d), using Section VI, Lemma 4 of [108], we have

$$\begin{aligned}
&\mathbb{E}\left[e^{\mathbb{i}tU_n^{\#}} \cdot e^{-(\rho_n \cdot n)^{-1}(t^2/2)\{\mathbb{E}[\sigma_w^2] + \frac{1}{n}\sum_{i=1}^n g_{\sigma;1}(X_i)\}}\right] \\
&= e^{-(\rho_n \cdot n)^{-1}(t^2/2)\cdot\mathbb{E}[\sigma_w^2]} \cdot \mathbb{E}\left[e^{\mathbb{i}t\sum_{i=1}^n g_1(X_i)/(\sqrt{n}\xi_1) - (\rho_n n)^{-1} t^2/(2n)\cdot\sum_{i=1}^n g_{\sigma;1}(X_i)}\right] \\
&= e^{-(\rho_n \cdot n)^{-1}(t^2/2)\cdot\mathbb{E}[\sigma_w^2]} \cdot \mathbb{E}\left[e^{\mathbb{i}t\sum_{i=1}^n g_1(X_i)/(\sqrt{n}\xi_1)} \cdot \left(1 + \widetilde{O}_p\left(\frac{t^2 \log^{1/2} n}{\rho_n n^{3/2}}\right)\right)\right] \\
&= e^{-(\rho_n \cdot n)^{-1}(t^2/2)\cdot\mathbb{E}[\sigma_w^2]} \cdot \mathbb{E}\left[e^{\mathbb{i}t\sum_{i=1}^n g_1(X_i)/(\sqrt{n}\xi_1)}\right] + O\Big(\frac{t^2 \log^{1/2} n}{\rho_n n^{3/2}}\Big) \\
&= e^{-(\rho_n \cdot n)^{-1}(t^2/2)\cdot\mathbb{E}[\sigma_w^2]} \cdot \mathbb{E}\left\{e^{-t^2/2} + O(n^{-1/2} t^3 e^{-t^2/2})\right\} + O\Big(\frac{t^2 \log^{1/2} n}{\rho_n n^{3/2}}\Big)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\left|\mathbb{E}\left[e^{\mathbb{i}tU_n^{\#}} \cdot e^{-(\rho_n \cdot n)^{-1} t^2/2\{\mathbb{E}[\sigma_w^2] + \frac{1}{n}\sum_{i=1}^n g_{\sigma;1}(X_i)\}}\right] - e^{-t^2/2}\right| \\
\leqslant & e^{-t^2/2}\left|e^{-C(\rho_n \cdot n)^{-1} t^2} - 1\right| + O(n^{-1/2} t^3 e^{-t^2/2}) + O\Big(\frac{t^2 \log^{1/2} n}{\rho_n n^{3/2}}\Big) \\
\leqslant & e^{-t^2/2} \cdot O((\rho_n \cdot n)^{-1} t^2) + O(n^{-1/2} t^3 e^{-t^2/2}) + O\Big(\frac{t^2 \log^{1/2} n}{\rho_n n^{3/2}}\Big)
\end{aligned}$$

Consequently,

$$\int_0^{\{C_0 \log(\rho_n \cdot n)\}^{1/2}} \left|\frac{\mathbb{E}[e^{\mathbb{i}t(U_n^{\#} + \widetilde{\Delta}_n)}] - e^{-t^2/2}}{t}\right| \mathrm{d}t = O((\rho_n \cdot n)^{-1}) \tag{8.103}$$

where we recall (8.102) to simplify notation. For the second part of the integral, we can reproduce the steps in the proof of Theorem 3.1 and obtain

$$\begin{aligned}
&\mathbb{E}\left[e^{\mathbb{i}tU_n^{\#}} \cdot e^{-(\rho_n \cdot n)^{-1}(t^2/2)\{\mathbb{E}[\sigma_w^2] + \frac{1}{n}\sum_{i=1}^n g_{\sigma;1}(X_i)\}}\right] \\
&= e^{-(\rho_n n)^{-1}(t^2/2)\mathbb{E}[\sigma_w^2]} \cdot \mathbb{E}\left[e^{\mathbb{i}tU_n^{\#}} \cdot e^{-(\rho_n n)^{-1} t^2/2\{\frac{1}{n}\sum_{i=1}^n g_{\sigma;1}(X_i)\}}\right] \\
&\leqslant (1 - C_1 \cdot t^2/n)^n \leqslant e^{-C_2 t^2} \leqslant e^{-C_2 \cdot C_0 \log(\rho_n \cdot n)} \leqslant (\rho_n \cdot n)^{-2}.
\end{aligned}$$

Therefore we have

$$\int_{\{C_0\log(\rho_n n)\}^{1/2}}^{(\rho_n n)\log^{-1/2} n}\left|\left[e^{\mathbb{i}tU_n^{\#}}\cdot e^{-(\rho_n\cdot n)^{-1}(t^2/2)\{\mathbb{E}[\sigma_w^2]+\frac{1}{n}\sum_{i=1}^n g_{\sigma;1}(X_i)\}}\right]\right|\cdot t^{-1}dt$$

$$\leqslant \int_{\{C_0\log(\rho_n n)\}^{1/2}}^{(\rho_n n)\log^{-1/2} n}\left|\left[e^{\mathbb{i}tU_n^{\#}}\cdot e^{-(\rho_n\cdot n)^{-1}(t^2/2)\{\mathbb{E}[\sigma_w^2]+\frac{1}{n}\sum_{i=1}^n g_{\sigma;1}(X_i)\}}\right]\right|dt$$

$$\leqslant \int_{\{C_0\log(\rho_n n)\}^{1/2}}^{(\rho_n n)\log^{-1/2} n}(\rho_n\cdot n)^{-2}\mathrm{d}t\leqslant(\rho_n\cdot n)^{-1}$$

Moreover, we choose $C_0\geqslant 4$ so that

$$\int_{\{C_0\log(\rho_n n)\}^{1/2}}^{(\rho_n n)\log^{-1/2} n}\frac{e^{-t^2/2}}{t}dt\leqslant\int_{\{C_0\log(\rho_n n)\}^{1/2}}^{(\rho_n n)\log^{-1/2} n}e^{-t^2/2}dt$$

$$\leqslant(\rho_n n)\log^{-1/2}(n)\cdot e^{-(C_0/2)\cdot\log(\rho_n n)}\leqslant\frac{\rho_n n}{(\rho_n n)^{C_0/2}}\leqslant\frac{1}{\rho_n n}.$$

Therefore, we have

$$\int_{\{C_0\log(\rho_n n)\}^{1/2}}^{(\rho_n n)\log^{-1/2} n}\left|\frac{\mathbb{E}[e^{\mathbb{i}t(U_n^{\#}+\tilde{\Delta}_n)}]-e^{-t^2/2}}{t}\right|\mathrm{d}t=O_p((\rho_n\cdot n)^{-1}) \quad (8.104)$$

Combining (8.103) and (8.104) proves (8.101).

- **Case 2:** $\rho_n\cdot n\log^{-1/2}n<\{C_0\log(\rho_n\cdot n)\}^{1/2}$. The proof in this case is even easier, since (8.103) remains valid and implies (8.101).

Plugging (8.101) back into the Esseen's smoothing lemma proves (8.100). Notice that the $\log^{1/2}n$ factor in the eventual error bound comes from the second term on the RHS of (8.53).

Next, reproducing the proof (8.55), we prove (8.99) by combining (8.100) and Lemma 3.1-(b).

Finally, the proof of (8.98) is done by combining (8.55) and Lemma 8.2. The proof of this part is exactly similar to the proof of (8.54). This completes the proof of the $O\left(\mathcal{M}(\rho_n,n;R)\right)$ bound.

Part II: proof of the $o(1)$ bound when $1\prec\rho_n n\preceq\log^{1/2}n$ (acyclic) or $1\prec\rho_n^{r/2}n\preceq\log^{1/2}n$ (cyclic).

The error bounds we derived in Part I of this proof focused on establishing finite sample error rates, and consequently need to bound the tail probability at the price of a $\log^{1/2}n$ factor multiplied on the error bound. Taking the acyclic motif setting as an example, to counter the log factor in the error bound, we also need to assume $\rho_n=\omega(n^{-1}\log^{1/2}n)$ rather than $\rho_n=\omega(n^{-1})$. For $\rho_n: n^{-1}\prec\rho_n\preceq n^{-1}\log^{1/2}n$, despite establishing an explicit finite-sample error bound is still possible, the formula and derivation are rather complicated. For cleanness of presentation, in this paper, we slightly lower the goal and only aim at deriving uniform consistency. Consequently, the proof can be done by slightly varying the proof of the first part of Theorem 3.4. In this proof, we do not need to show an explicit error rate, so we do not need "$\widetilde{O}_p$" any more, and "o_p" would suffice for our purpose. For the convenience of narration, we define

$$\widetilde{\mathcal{M}}(\rho_n,n;R):=\begin{cases}(\rho_n\cdot n)^{-1/2}+n^{-1}\cdot\log^{3/2}n, & \text{For acyclic } R\\ (\rho_n^{r/2}\cdot n)^{-1/2}+n^{-1}\cdot\log^{3/2}n, & \text{For cyclic } R\end{cases}$$

We first present a variant of Lemma 3.1.

LEMMA 8.4. *Under the conditions of Theorem 3.4, we have the following results:*

(a) *Identical to Lemma 3.1-(a).*

(b) *We have*

$$\widehat{\Delta}_n := \frac{\widehat{U}_n - U_n}{\sigma_n} = \breve{\Delta}_n + \breve{R}_n$$

where $\breve{\Delta}_n$ and $\breve{R}_n$ satisfy

$$\breve{R}_n = o_p(\widetilde{\mathcal{M}}(\rho_n, n; R)) \tag{8.105}$$

and the original (3.12) in Lemma 3.1-(b) holds for $\breve{\Delta}_n$, where the definition and asymptotic order of σ_w is identical to that in Lemma 3.1,

(c) $\widehat{\delta}_n = o_p(\widetilde{\mathcal{M}}(\rho_n, n; R))$,

(d) *Identical to Lemma 3.1-(d).*

PROOF OF LEMMA 8.4. The proof of this lemma can be obtained by slightly varying the proof of Lemma 3.1.

(a) (No additional proof needed.)

(b) The only change we need to make to the proof of Lemma 3.1-(b) to make it a valid proof here is to replace (8.18) by the following concentration inequality:

(8.106)

$$\begin{aligned}
&\mathbb{P}\left(\breve{R}_n := \frac{\mathring{R}}{\binom{n}{r}\cdot\sigma_n} \geqslant C\cdot\widetilde{\mathcal{M}}(\rho_n, n; R)\right)\\
\preceq &\begin{cases} \max\left\{\exp\left(-\frac{((\rho_n\cdot n)^{-1}\cdot(\rho_n\cdot n)^{1/2})^2}{(\rho_n\cdot n)^{-2}}\right), \exp\left(-\frac{(\rho_n\cdot n)^{-1}\cdot(\rho_n\cdot n)^{1/2}}{(\rho_n\cdot n)^{-1}\cdot n^{-1/2}}\right)\right\}, & \text{for acyclic } R;\\ \max\left\{\exp\left(-\frac{((\rho_n^{-r/2}\cdot n^{-1})\cdot(\rho_n^{r/2}\cdot n)^{1/2})^2}{(\rho_n^{-r/2}\cdot n^{-1})^2}\right), \exp\left(-\frac{(\rho_n^{-r/2}\cdot n^{-1})\cdot(\rho_n^{r/2}\cdot n)^{1/2}}{\rho_n^{-3} n^{-5/2}}\right)\right\}, & \text{for cyclic } R;\end{cases}\\
=&o(1)
\end{aligned}$$

The proof of this part is completed.

(c) We only need to change how we use Theorem 3 of Schudy and Sviridenko [116] in (8.42), into the following way

$$\sum_{\substack{\text{All possible } (v,p):\\ v\geqslant 2, p\geqslant 3}} \widehat{\Delta}^{(i;v,p)} = o_p(\rho_n^s \cdot n^{r-1}\cdot\widetilde{\mathcal{M}}(\rho_n, n; R))$$

and for the rest of the proof of Lemma 3.1-(c), replace every remainder term in the format of "$\widetilde{O}_p(\cdots\times\mathcal{M}(\rho_n, n; R))$" by "$o_p(\cdots\times\widetilde{\mathcal{M}}(\rho_n, n; R))$". This completes the proof.

(d) (No additional proof needed.)

□

Now we return to the proof of the second part of Theorem 3.4. The proof is completed by slightly varying (8.97) in the proof of the first part of this theorem by

$$\widehat{T}_n = U_n^{\#} + \widehat{\Delta}_n(1 + o_p(\widetilde{\mathcal{M}}(\rho_n, n; R)))) + o_p(\widetilde{\mathcal{M}}(\rho_n, n; R)))$$

Then recall the definition of "o_p" and apply Lemma 2 of Maesono [96] (setting $T = \widehat{T}_n$, $\widetilde{T} = U_n^{\#}$ and $\alpha = \widetilde{\mathcal{M}}(\rho_n, n; R)$ and $H(x) = \Phi(x)$). We have

$$\begin{aligned}
\|F_{\widehat{T}_n}(u) - \Phi(u)\|_\infty \leqslant &\|F_{U_n^{\#}}(u) - \Phi(u)\|_\infty + \mathbb{P}\left\{|\widehat{T}_n - U_n^{\#}| \geqslant \widetilde{\mathcal{M}}(\rho_n, n; R)\right\}\\
&+ O(\widetilde{\mathcal{M}}(\rho_n, n; R)) = o(1) + o(1) + o(1) \to 0
\end{aligned}$$

This completes the proof of the second part and thus the proof of the error bound of the population version Edgeworth expansion in Theorem 3.4.

Next, we prove the error bound for the empirical version Edgeworth expansion in Theorem 3.4. Similar to the proof of the population version, we discuss two cases.

Part I: the proof of the $\widetilde{O}_p\big(\mathcal{M}(\rho_n, n; R)\big)$ bound when $\rho_n n \succ \log^{1/2} n$ (acyclic) or $\rho_n^{r/2} n \succ \log^{1/2} n$ (cyclic) is easily done by citing the following intermediate results from the proof of Theorem 3.2.

$$\widehat{\xi}_1^3 - \xi_1^3 = \widetilde{O}_p(\rho_n^{3s} \cdot n^{-1/2} \log^{1/2} n), \tag{8.107}$$

$$\big|\widehat{\mathbb{E}}[g_1^3(X_1)] - \mathbb{E}[g_1^3(X_1)]\big| = \widetilde{O}_p(\rho_n^{3s-1/2} \cdot n^{-1/2} \log^{1/2} n), \tag{8.108}$$

$$\begin{aligned}&\Bigg|\binom{n}{2}^{-1} \sum_{1\leqslant i<j\leqslant n} \widehat{g}_1(X_1)\widehat{g}_1(X_2)\widehat{g}_2(X_1,X_2) - \mathbb{E}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)]\Bigg|\\ &= \widetilde{O}_p(\rho_n^{3s-1/2} \cdot n^{-1/2} \log^{1/2} n).\end{aligned} \tag{8.109}$$

The proof of this part is then instantly done by combining these results with the statement about the population Edgeworth expansion in the sparse case that we just proved above.

Part II: proof of the $o_p(1)$ bound when $1 \prec \rho_n n \le \log^{1/2} n$ (acyclic) or $1 \prec \rho_n^{r/2} n \le \log^{1/2} n$. To prove for this regime, we only need to slightly vary the proof of Theorem 3.2. Set a series $\varrho_n \to \infty$ as follows:

$$\varrho_n := \begin{cases} \rho_n \cdot n, & \text{for acyclic } R,\\ \rho_n^{r/2} \cdot n, & \text{for cyclic } R. \end{cases} \tag{8.110}$$

By replacing the $\log^{1/2} n$ factor in all the "u" values that we set in Theorem 8.1 by ϱ_n, where we apply it in the proof of Lemma 3.1-(c),(d) and in the proof of Theorem 3.2, we establish the following analogous intermediate results:

$$\widehat{\xi}_1^3 - \xi_1^3 = O_p(\rho_n^{3s} \cdot n^{-1/2} \cdot \varrho_n) = o_p(\rho_n^{3s}) \tag{8.111}$$

$$\big|\widehat{\mathbb{E}}[g_1^3(X_1)] - \mathbb{E}[g_1^3(X_1)]\big| = O_p(\rho_n^{3s-1/2} \cdot n^{-1/2} \cdot \varrho_n) = o_p(\rho_n^{3s-1/2}) \tag{8.112}$$

$$\begin{aligned}&\Bigg|\binom{n}{2}^{-1} \sum_{1\leqslant i<j\leqslant n} \widehat{g}_1(X_1)\widehat{g}_1(X_2)\widehat{g}_2(X_1,X_2) - \mathbb{E}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)]\Bigg|\\ &= O_p(\rho_n^{3s-1/2} \cdot n^{-1/2} \cdot \varrho_n) = o_p(\rho_n^{3s-1/2})\end{aligned} \tag{8.113}$$

This implies $\|G_n(u) - \widehat{G}_n(u)\|_\infty = o_p\big((\rho_n n)^{-1/2}\big) = o_p(1)$ (thus immediately completes the proof of Part II) by simply reproducing the rest of the proof of Theorem 3.2. The proof of the entire Theorem 3.4 is now complete.

□

8.6. *Proof of Theorem 4.1*. We will mainly prove for the node sub-sampling network bootstrap scheme [17], and the corresponding conclusion for the re-sampling scheme can be obtained easily by slightly varying the proof for sub-sampling. Conditioned on A, since the sub-sampling objects in network models are nodes rather than latent variables X_j's[3],

[3] In other words, X_j's in the bootstrap procedure are deemed fixed.

we change the notation for simplicity. Define $\mathcal{V}_\star = \{1 \leqslant v_1 < v_2 < \cdots < v_{n^*} \leqslant n\}$ to be uniformly sampled from all size-n^* subsets of $[n]$. That is,

$$\mathbb{P}\Big(\mathcal{V}_\star = \{i_1, \cdots, i_{n^*}\}\Big) = \frac{1}{\binom{n}{n^*}} \quad \forall 1 \leqslant i_1 < \cdots < i_{n^*} \leqslant n.$$

Define the bootstrap expectation $\mathbb{E}^*$ to be taken with respect to the randomness of $\mathcal{V}_\star$. The sub-sampling bootstrap sample network moment $\widehat{U}^b_{n^*}$ calculated from the sub-network $A_{\mathcal{V}_\star,\mathcal{V}_\star}$ calculated according to [17] is

$$\widehat{U}^b_{n^*} = \frac{1}{\binom{n^*}{r}} \sum_{i_1<\cdots<i_r\subset\mathcal{V}_\star} h(A_{i_1,i_2,\cdots,i_r}).$$

To emphasize that the randomness in this bootstrap setting is solely due to $\mathcal{V}_*$ and simplify notation, we define $\widehat{g}^b_1(v_1)$, taking the argument v_1 rather than X_{v_1}, as follows

(8.114)
$$\widehat{g}^b_1(v_1) := \frac{n-1}{n-n^\star}\left\{\frac{1}{\binom{n^*-1}{r-1}}\mathbb{E}^*\Big[\sum_{i_1,\cdots,i_{r-1}\subset\mathcal{V}_\star\backslash v_1} h(A_{v_1,i_1,\cdots,i_{r-1}})|v_1\Big] - \widehat{U}_n\right\}$$

$$\widehat{g}^b_2(v_1,v_2) := \frac{n-3}{n-n^*-1}\Bigg(\frac{n-2}{n-n^*}\Big\{\mathbb{E}^*\Big[\frac{1}{\binom{n^*-2}{r-2}}\sum_{i_1,\cdots,i_{r-2}\subset\mathcal{V}_\star\backslash\{v_1,v_2\}} h(A_{v_1,v_2,i_1,\cdots,i_{r-2}})\Big|v_1,v_2\Big]$$

(8.115)
$$-\widehat{U}_n\Big\} - \widehat{g}^b_1(v_1) - \widehat{g}^b_1(v_2)\Bigg)$$

where the finite population correction term $(n-1)/(n-n^*)$ comes from [23, (1.2)]. where again the finite population correction term $(n-3)/(n-n^*-1)$ is due to [23, (1.3)]. Recall that $\widehat{S}^*_{n^*}$ is a jackknife estimator of $\mathrm{Var}^*\Big(\widehat{U}^b_{n^*}|A\Big)$ and that the bootstrap test statistic as

$$\widehat{T}^*_{n^*} = \frac{\widehat{U}^b_{n^*} - \widehat{U}_n}{\widehat{S}^*_{n^*}} \tag{8.116}$$

By our proof of Theorem 3.3, the difference between a jackknife estimator and an estimator based on ξ^*_1 is ignorable, and we are free to choose either. Here we use the jackknife estimator in order to better connect with Bloznelis [23]. To start, we check that $\mathbb{E}^*[\widehat{U}^b_{n^*}] = \widehat{U}_n$ where the expectation is taken with respect to the randomness of $\mathcal{V}_\star$, so that (8.116) is an valid studentization of the U-statistic. To see this, notice that

$$\mathbb{E}^*[\widehat{U}^b_{n^*}] = \frac{1}{\binom{n}{n^*}}\sum_{\mathcal{V}_\star\subset[n]}\widehat{U}^b_{n^*} = \frac{1}{\binom{n}{n^*}}\frac{1}{\binom{n^*}{r}}\sum_{\mathcal{V}_\star\subset[n]}\sum_{i_1<\cdots<i_r\subset\mathcal{V}_\star} h(A_{i_1,i_2,\cdots,i_r}).$$

On the RHS, each summand $h(A_{i_1,\cdots,i_r})$ appears $\binom{n-r}{n^*-r}$ times. Therefore,

$$\begin{aligned}\sum_{\mathcal{V}_\star\subset[n]}\sum_{i_1<\cdots<i_r\subset\mathcal{V}_\star} h(A_{i_1,i_2,\cdots,i_r}) &= \binom{n-r}{n^*-r}\sum_{1\leqslant i_1<\cdots<i_r\leqslant n} h(A_{i_1,\cdots,i_r}) \\ &= \binom{n-r}{n^*-r}\binom{n}{r}\widehat{U}_n.\end{aligned}$$

As a result,

$$\mathbb{E}^*[\widehat{U}^b_{n*}] = \frac{1}{\binom{n}{n*}} \frac{1}{\binom{n*}{r}} \sum_{\mathcal{V}_\star \subset [n]} \sum_{i_1 < \cdots < i_r \subset \mathcal{V}_\star} h(A_{i_1,i_2,\cdots,i_r})$$

$$= \frac{1}{\binom{n}{n*}} \frac{1}{\binom{n*}{r}} \binom{n-r}{n^*-r} \binom{n}{r} \cdot \widehat{U}_n = \widehat{U}_n$$

To investigate the distribution of $\widehat{T}^*_{n*}$ under the finite-population sampling obeying $\mathcal{V}_\star$, we define the bootstrap Edgeworth expansion by

$$G^*_{n*}(x) := \Phi(x) + \frac{\varphi(x)}{\sqrt{n^*(1-n^*/n)} \cdot (\xi^*_1)^3} \cdot \left\{ \frac{2x^2+1}{6} \cdot \mathbb{E}^* \left\{ \widehat{g}^b_1(v_1) \right\}^3 \right.$$

$$\left. + \frac{r-1}{2} \cdot \left(x^2+1\right) \mathbb{E}^*[\widehat{g}^b_1(v_1)\widehat{g}^b_1(v_2)\widehat{g}^b_2(v_1,v_2)] \right\} \tag{8.117}$$

where recall the definitions of $\widehat{g}^b_1(\cdot), \widehat{g}^b_2(\cdot,\cdot)$ from (8.114) and (8.115), respectively. Here, $(\xi^*_1)^2 := \mathrm{Var}^*(\widehat{g}^b_1(v_1)|A) = \mathbb{E}^*[(\widehat{g}^b_1(v_1))^2]$.

Next, we are going to apply Theorem 1 of [23]. The Cramer's condition (1.11) in Theorem 1 in [23] is different from the conventional version, and we need to check that it indeed holds in our setting. Specifically, in our setting, it suffices to prove that there exists a positive sequence $\{t_n\} \to \infty$ and a universal constant $M_1 : 0 < M_1 < 1$, such that

$$\mathbb{P}\left(\sup_{t \in (0,t_n)} \left| n^{-1} \sum_{j=1}^n e^{\mathrm{i} t_n \widehat{g}_1(X_j)/\widehat{\xi}_1} \right| \leqslant M_1 < 1 \right) \xrightarrow{p} 1$$

because our eventual bounds are O_p bounds, and in the proof we can choose to discuss only events that will happen with high probability. Recall from the proof of Theorem 3.2 that we have shown the following facts

$$|\widehat{g}_1(X_i) - g_1(X_i)| = \widetilde{O}_p(\rho_n^{s-1/2} \cdot n^{-1/2} \log^{1/2} n)$$

$$|\widehat{\xi}_1 - \xi_1| = \widetilde{O}_p(\rho_n^{s} \cdot n^{-1/2} \log^{1/2} n)$$

and the simple fact that $\xi_1 \asymp \rho_n^s$ a.s. Therefore, we have

$$|\widehat{g}_1(X_j)/\widehat{\xi}_1 - g_1(X_j)/\xi_1| = \widetilde{O}_p(\rho_n^{-1/2} \cdot n^{-1/2} \log^{1/2} n)$$

Recall that our assumption implies $\rho_n n \to \infty$ throughout this paper (regardless of R shapes, all assumptions we made imply this). Choosing $t_n = (\rho_n \cdot n)^{1/4}$, we have

$$\sup_{t \in (0,t_n)} \left| \frac{1}{n} \sum_{j=1}^n e^{\mathrm{i} t g_1(X_j)/\xi_1} - \frac{1}{n} \sum_{j=1}^n e^{\mathrm{i} t \widehat{g}_1(X_j)/\widehat{\xi}_1} \right|$$

$$\leqslant \sup_{t \in (0,t_n)} t \cdot \max_{1 \leqslant j \leqslant n} \left| g_1(X_j)/\xi_1 - \widehat{g}_1(X_j)/\widehat{\xi}_1 \right| \cdot e^{t \cdot \left| g_1(X_j)/\xi_1 - \widehat{g}_1(X_j)/\widehat{\xi}_1 \right|}$$

$$(\text{w.p. } 1 - Cn^{-1}) \leqslant \sup_{t \in (0,t_n)} t(\rho_n \cdot n)^{-1/2} \log^{1/2} n \cdot e^{t(\rho_n \cdot n)^{-1/2} \log^{1/2} n} \preceq t_n(\rho_n \cdot n)^{-1/2}$$

under the specified sparsity conditions.

It suffices to bound $\sup_{t\in(0,t_n)}\left|n^{-1}\sum_{j=1}^n e^{\mathbb{i}tg_1(X_j)/\xi_1}\right|$. For every given $t\in\mathcal{T}_n:=\{k/n:k\in\mathbb{N},k/n\leqslant t_n\}$, by Bernstein's inequality, we have

$$\mathbb{P}\left(\left|n^{-1}\sum_{j=1}^n e^{\mathbb{i}tg_1(X_j)/\xi_1}-\mathbb{E}\left[e^{\mathbb{i}tg_1(X_1)/\xi_1}\right]\right|>\epsilon\right)\leqslant 2e^{-Cn\epsilon^2}$$

Therefore, setting $M_1:=\limsup_{t\to\infty}\left|\mathbb{E}\left[e^{\mathbb{i}tg_1(X_1)/\xi_1}\right]\right|$, by the Cramer's condition we assumed in Theorem 4.1, we have $M_1\in(0,1)$ and $(1+M_1)/2\in(0,1)$. Therefore

$$\mathbb{P}\left(\sup_{t\in\mathcal{T}_n}\left|n^{-1}\sum_{j=1}^n e^{\mathbb{i}tg_1(X_j)/\xi_1}\right|>(1+M_1)/2\right)\leqslant|\mathcal{T}_n|\cdot 2e^{-C_3n(M_1/2)^2}\leqslant e^{-C_4n}$$

for some universal constants $C_3,C_4>0$. Now noticing that for any $t\in(0,t_n)$, let t' be the best approximation to t in $\mathcal{T}_n$, we have

$$\sup_{t\in(0,t)}\left|\frac{1}{n}\sum_{j=1}^n e^{\mathbb{i}tg_1(X_j)/\xi_1}-\frac{1}{n}\sum_{j=1}^n e^{\mathbb{i}t'g_1(X_j)/\xi_1}\right|$$

$$\text{(w.h.p.)}\leqslant|t-t'|(\rho_n\cdot n)^{-1/2}\cdot e^{|t-t'|(\rho_n\cdot n)^{-1/2}}\preceq t_n\cdot(\rho_n\cdot n)^{-1/2}\to 0$$

The verification that our ordinary Cramer's condition implies the sample version in [23] is thus finished.

By Theorem 1 of [23], the sampling distribution of $\widehat{T}^*_{n^*}$ by node sub-sampling enjoys the following uniform bound

$$\left\|F_{\widehat{T}^*_n}(u)-G^*_{n^*}(u)\right\|_\infty=o_p((n^*)^{-1/2}) \tag{8.118}$$

It then suffices to establish the connection between $G^*_{n^*}(u)$ and $\widehat{G}_{n^*(1-n^*/n)}(u)$. The proof strategy is to show that (8.117) can be transcribed, with $\mathbb{E}^*$ replaced by $\widehat{\mathbb{E}}$'s and $\widehat{g}_1^b(v_1),\widehat{g}_2^b(v_1,v_2)$ replaced with $\widehat{g}_1(X_1),\widehat{g}_2(X_1,X_2)$, respectively. Then the comparison of the Edgeworth coefficients in $G^*_{n^*}(u)$ and $\widehat{G}_{n^*(1-n^*/n)}(u)$ would complete the proof. To proceed, now we focus on analyzing the core quantities $\widehat{g}_1^b(v_1)$ and $\widehat{g}_2^b(v_1,v_2)$. For $\widehat{g}_1^b(v_1)$, since conditioning on $v_1\in\mathcal{V}_\star$, the rest indexes $\{v_2,\cdots,v_{n^*}\}$ are uniformly sampled from $\{\{i_1,\cdots,i_{n^*-1}\}\subset[n]\backslash v_1\}$, we have

$$\begin{aligned}&\frac{1}{\binom{n^*-1}{r-1}}\cdot\mathbb{E}^*\left[\sum_{i_1,\cdots,i_{r-1}\subset\mathcal{V}_\star\backslash v_1}h(A_{v_1,i_1,\cdots,i_{r-1}})\middle|v_1\right]\\
&=\frac{1}{\binom{n^*-1}{r-1}}\frac{1}{\binom{n-1}{n^*-1}}\sum_{\mathcal{V}_\star\subset[n]:v_1\in\mathcal{V}_\star}\ \sum_{i_1,\cdots,i_{r-1}\in\mathcal{V}_\star\backslash v_1}h(A_{v_1,i_1,\cdots,i_{r-1}})\\
\text{(By (8.26))}&=\frac{1}{\binom{n^*-1}{r-1}}\frac{1}{\binom{n-1}{n^*-1}}\binom{n-r}{n^*-r}\binom{n-1}{r-1}\cdot\widehat{a}_{v_1}=\widehat{a}_{v_1}.\end{aligned}$$

where in the second equality, we noticed that each $h(A_{v_1,i_1,\cdots,i_{r-1}})$ appears $\binom{n-r}{n^*-r}$ times in the first line. Therefore,

$$\widehat{g}_1^b(v_1)=\frac{n-1}{n-n^*}[\widehat{a}_{v_1}-\widehat{U}_n]=\frac{n-1}{n-n^*}\cdot\widehat{g}_1(X_{v_1}) \tag{8.119}$$

where $\widehat{g}_1(X_{v_1})$ appeared (in "$\widehat{\mathbb{E}}$" terms) in Theorem 3.2. Then we have

$$\mathbb{E}^*\left[\{\widehat{g}_1^b(v_1)\}^3\right] = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{n-1}{n-n^*}\right)^3(\widehat{a}_i - \widehat{U}_n)^3 = \left(\frac{n-1}{n-n^*}\right)^3\widehat{\mathbb{E}}[g_1^3(X_1)]$$

$$(\xi_1^*)^2 = \mathrm{Var}^*(\widehat{g}_1^b(v_1)|A) = \mathbb{E}^*[(\widehat{g}_1^b(v_1))^2] = \frac{(n-1)^2}{(n-n^*)^2}\cdot\widehat{\xi}_1^2$$

where the definitions of $\widehat{\xi}_1$ and $\widehat{\mathbb{E}}g_1^3(X_1)$ can also be recalled by reviewing Theorem 3.2. Now we turn to analyzing $\mathbb{E}^*[\{\widehat{g}_1^b(v_1)\widehat{g}_1^b(v_2)\widehat{g}_2^b(v_1,v_2)\}]$. The main part of the definition of $\widehat{g}_2^b(v_1,v_2)$ can be re-expressed as follows

$$\begin{aligned}
&\mathbb{E}^*\left[\frac{1}{\binom{n^*-2}{r-2}}\sum_{i_1,\cdots,i_{r-2}\subset\mathcal{V}_\star\setminus\{v_1,v_2\}}h(A_{v_1,v_2,i_1,\cdots,i_{r-2}})\Big|v_1,v_2\right]\\
=&\frac{1}{\binom{n-2}{n^*-2}}\frac{1}{\binom{n^*-2}{r-2}}\sum_{\mathcal{V}_\star\subset[n]:v_1,v_2\in\mathcal{V}_\star}\sum_{i_1,\cdots,i_{r-2}\subset\mathcal{V}_\star\setminus\{v_1,v_2\}}h(A_{v_1,v_2,i_1,\cdots,i_{r-2}})\\
=&\frac{1}{\binom{n-2}{n^*-2}}\frac{1}{\binom{n^*-2}{r-2}}\binom{n-2}{r-2}\binom{n-r}{n^*-r}\widehat{a}_{v_1v_2} = \widehat{a}_{v_1v_2}
\end{aligned}$$

where we recall the definition of $\widehat{a}_{ij}$ from (8.82). Combining this with (8.119), we have

$$\begin{aligned}
\widehat{g}_2^b(v_1,v_2) =&\frac{n-3}{(n-n^*-1)}\left[\frac{n-2}{n-n^*}(\widehat{a}_{v_1v_2}-\widehat{U}_n) - \frac{n-1}{n-n^*}(\widehat{a}_{v_1}-\widehat{U}_n) - \frac{n-1}{n-n^*}(\widehat{a}_{v_2}-\widehat{U}_n)\right]\\
=&\frac{(n-3)(n-1)}{(n-n^*-1)(n-n*)}\left[(\widehat{a}_{v_1v_2}-\widehat{U}_n) - (\widehat{a}_{v_1}-\widehat{U}_n) - (\widehat{a}_{v_2}-\widehat{U}_n)\right]\\
&-\frac{(n-3)}{(n-n^*-1)(n-n*)}(\widehat{a}_{v_1v_2}-\widehat{U}_n).
\end{aligned}$$

Then we have

$$\begin{aligned}
&\mathbb{E}^*[\widehat{g}_1^b(v_1)\widehat{g}_1^b(v_2)\widehat{g}_2^b(v_1,v_2)] = \frac{1}{\binom{n}{2}}\sum_{1\leqslant v_1<v_2\leqslant n}\widehat{g}_1^b(v_1)\widehat{g}_1^b(v_2)\widehat{g}_2^b(v_1,v_2)\\
=&\frac{(n-3)(n-1)^3}{(n-n^*-1)(n-n*)^3}\widehat{\mathbb{E}}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)]\\
&-\frac{(n-3)(n-1)^2}{(n-n^*-1)(n-n^*)^3}\cdot\frac{1}{\binom{n}{2}}\sum_{1\leqslant i<j\leqslant n}\widehat{g}_1(X_1)\widehat{g}_1(X_2)[\widehat{g}_2(X_1,X_2)+\widehat{g}_1(X_1)+\widehat{g}_1(X_2)]\\
=&\frac{(n-3)(n-1)^3}{(n-n^*-1)(n-n*)^3}\widehat{\mathbb{E}}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)] + \widetilde{O}_p\left(\frac{(n-3)(n-1)^2}{(n-n^*-1)(n-n^*)^3}\cdot\rho_n^{3s-1}\log^{1/2}n\right)
\end{aligned}$$

where in the last line, we used that, $\widehat{g}_1(X_1)\overset{p}{\asymp}\rho_n^s, \widehat{g}_2(X_1,X_2)\overset{p}{\asymp}\rho_n^{s-1}$ with probability at least $1-O(n^{-1})$ by the proof of Theorem 3.2. Define $\alpha_{n*} = (n-1)/(n-n^*)$. Now we can rewrite (8.117) as follows

$$\begin{aligned}
G_{n^*}^*(x) =&\Phi(x) + \frac{\varphi(x)}{\sqrt{n^*(1-n^*/n)}\cdot\alpha_{n^*}^3\widehat{\xi}_1^3}\left\{\frac{2x^2+1}{6}\cdot\alpha_{n^*}^3\widehat{\mathbb{E}}[g_1^3(X_1)]\right.\\
&+\frac{r-1}{2}\cdot\alpha_{n^*}^3(x^2+1)\widehat{\mathbb{E}}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)]
\end{aligned}$$

$$-\frac{n-3}{(n-n^*-1)(n-n^*)}\frac{r-1}{2}\cdot\alpha_{n^*}^2(x^2+1)\cdot\widetilde{O}_p(\rho_n^{3s-1}\log^{1/2}n)\Big\}$$

$$=\widehat{G}_{n^*(1-n^*/n)}(u)+\widetilde{O}_p\left\{\frac{\log^{1/2}n}{\sqrt{n^*(1-n^*/n)(n-n^*)\rho_n}}\right\}$$

where recall that $\widehat{G}_n(u)$ was defined Theoem 3.2. Finally, we have

$$\left\|G_{n^*}^*(u)-\widehat{G}_{n^*(1-n^*/n)}(u)\right\|_\infty=\widetilde{O}_p\left\{\frac{\log^{1/2}(n)}{\sqrt{n^*(1-n^*/n)(n-n^*)\rho_n}}\right\}$$

where the last equation is due to $\rho_n=\omega(n^{-1/r})$ and $n-n^*\asymp n$. Combining this with Theorem 3.1 and Theorem 3.2, by a triangular inequality, we have

$$\left\|F_{\widehat{T}_n^*}(u)-F_{\widehat{T}_{n^*(1-n^*/n)}}(u)\right\|_\infty=o_p((n^*)^{-1/2}). \tag{8.120}$$

This completes the proof of Theorem 4.1 for sub-sampling, since the uniform convergence rate of the Edgeworth expansion is governed by the worst convergence rate of its coefficient terms.

Now we discuss the re-sampling scheme. Sampling $\{v_1,\cdots,v_{n^*}\}$ with replacement from a finite population $[n]$ is equivalent to sampling without replacement from a population in which each of $[n]$ are repeated infinite many times with the same infinite cardinality such that a uniform sampler will still take each of $[n]$ with equal probabilities. This amounts to set the "n" in Bloznelis [23] to "$n=\infty$"[4]. Notice, however, the "n" in [23] should not be confused with our network size n in the expressions of ξ_1^*, $\mathbb{E}^*[\{\widehat{g}_1^b(v_1)\}^3]$ and $\mathbb{E}^*[\widehat{g}_1^b(v_1)\widehat{g}_1^b(v_2)\widehat{g}_2^b(v_1,v_2)]$. Therefore, the re-sampling bootstrap Edgeworth expansion is the following slight-modification of (8.117):

$$\begin{aligned}G_{n^*}^*(x):=\Phi(x)+\frac{\varphi(x)}{\sqrt{n^*}\cdot(\xi_1^*)^3}\cdot\Bigg\{&\frac{2x^2+1}{6}\cdot\mathbb{E}^*\left\{\widehat{g}_1^b(v_1)\right\}^3\\ &+\frac{r-1}{2}\cdot\left(x^2+1\right)\mathbb{E}^*[\widehat{g}_1^b(v_1)\widehat{g}_1^b(v_2)\widehat{g}_2^b(v_1,v_2)]\Bigg\}\end{aligned} \tag{8.121}$$

The rest of the proof is exactly similar to that for sub-sampling and thus will be omitted. The proof of Theorem 4.1 is completed.

PROOF OF THEOREM 4.2. The key to this proof is to establish the local monotonicity of the function $G_n(\cdot)$. The local curvature of G_n is handier to use than that of $F_{\widehat{T}_n}$, because the distribution of $\widehat{T}_n$ may not be exactly continuous, and the classical result $F_Z(Z)\sim$ Uniform$[0,1]$ (thus $\mathbb{P}(F_Z(Z)\leqslant u)=u$ for any $u\in[0,1]$) for a continuous random variable Z does not necessarily apply. On the other hand, by construction, G_n is always smooth.

Now notice that not only $G_n(\cdot)$ uniformly converges to the $N(0,1)$ CDF $\Phi(\cdot)$, but further, these two functions are both smooth and $\sup_u|G_n'(u)-\Phi'(u)|\to 0$ (while the CDF $F_{\widehat{T}_n}(\cdot)$ is not necessarily continuous). Therefore, there exists a large enough constant n_0 and small constants $\epsilon_0>0,\delta_0>0$, such that the following two properties hold simultaneously

[4]Here, we clarify that the "n" in "$n=\infty$" should be understood as the size of the finite population for bootstrapping, among the notation system of [23], not the "n" in most of this paper as the network size.

(i). For all $n \geqslant n_0$, we have $G_n(u) \geqslant \alpha/2 + \epsilon_0$ for all $u \geqslant z_{\alpha/2} + \delta_0$; and $G_n(u) \leqslant \alpha/2 - \epsilon_0$ for all $u \leqslant z_{\alpha/2} - \delta_0$

(ii). For all $n \geqslant n_0$, on the interval $u \in [z_{\alpha/2} - \delta_0, z_{\alpha/2} + \delta_0]$, we have $0 < C_{\alpha/2} \leqslant G_n'(u) \leqslant D_{\alpha/2}$ and constants $C_{\alpha/2}, D_{\alpha/2}$ only depend on α.

Properties (ii) implies that G_n is strictly monotone and thus invertible in $[z_{\alpha/2} - \delta_0, z_{\alpha/2} + \delta_0]$, and specifically, $G_n^{-1}(\alpha/2)$ is well-defined. Then by (i), we have $G_n^{-1}(\alpha/2) \in [z_{\alpha/2} - \delta_0, z_{\alpha/2} + \delta_0]$, and by (ii), we know that $G_n^{-1}(u')$ is also Lipschitz on $u' \in [G_n(z_{\alpha/2} - \delta_0), G_n(z_{\alpha/2} + \delta_0)]$.

Now we are ready to begin the main proof for Type-I error rate. We have

$$\begin{aligned}
\text{Type-I error rate} :=& \mathbb{P}_{H_0}\left(2 \cdot \min\left\{\widehat{G}_n(\widehat{T}_n), 1 - \widehat{G}_n(\widehat{T}_n)\right\} < \alpha\right) \\
(\alpha \text{ is small}) =& \mathbb{E}_{H_0}\left[\mathbb{1}_{[\widehat{G}_n(\widehat{T}_n) \leqslant \alpha/2]} + \mathbb{1}_{[\widehat{G}_n(\widehat{T}_n) > 1 - \alpha/2]}\right] \\
(\text{Theorems 3.1 + 3.2}) =& \mathbb{E}_{H_0}\left[\mathbb{1}_{[G_n(\widehat{T}_n) \leqslant \alpha/2]} + \mathbb{1}_{[G_n(\widehat{T}_n) > 1 - \alpha/2]}\right] \\
& + \mathbb{E}_{H_0}\left[\mathbb{1}_{[\widehat{G}_n(\widehat{T}_n) \leqslant \alpha/2]} - \mathbb{1}_{[G_n(\widehat{T}_n) \leqslant \alpha/2]}\right] \\
& + \mathbb{E}_{H_0}\left[\mathbb{1}_{[\widehat{G}_n(\widehat{T}_n) > 1 - \alpha/2]} - \mathbb{1}_{[G_n(\widehat{T}_n) > 1 - \alpha/2]}\right] \\
=& \mathbb{P}_{H_0}\left(G_n(\widehat{T}_n) \leqslant \alpha/2\right) + \mathbb{P}_{H_0}\left(G_n(\widehat{T}_n) \geqslant 1 - \alpha/2\right) \\
& + O\left(\mathcal{M}(\rho_n, n; R)\right),
\end{aligned} \tag{8.122}$$

where the last equality is due to (recall from the proof of Theorem 3.2 that $\|\widehat{G}_n(x) - G_n(x)\|_\infty = \widetilde{O}_p(\rho_n^{-1} n^{-1})$)

$$\begin{aligned}
& \mathbb{E}_{H_0}\left|\mathbb{1}_{[\widehat{G}_n(\widehat{T}_n) \leqslant \alpha/2]} - \mathbb{1}_{[G_n(\widehat{T}_n) \leqslant \alpha/2]}\right| \\
&\qquad = \mathbb{P}_{H_0}\big(\widehat{G}_n(\widehat{T}_n) \leqslant \alpha/2, G_n(\widehat{T}_n) > \alpha/2\big) + \mathbb{P}_{H_0}\big(\widehat{G}_n(\widehat{T}_n) > \alpha/2, G_n(\widehat{T}_n) \leqslant \alpha/2\big) \\
&\qquad = \mathbb{P}_{H_0}\big(G_n(\widehat{T}_n) \leqslant \alpha/2 + O(\rho_n^{-1} n^{-1}), G_n(\widehat{T}_n) > \alpha/2\big) \\
&\qquad\quad + \mathbb{P}_{H_0}\big(G_n(\widehat{T}_n) > \alpha/2 - O(\rho_n^{-1} n^{-1}), G_n(\widehat{T}_n) \leqslant \alpha/2\big) + O(n^{-1}) \\
(\text{Invertibility of } G_n(\cdot)) &\qquad = \mathbb{P}_{H_0}\left(G_n^{-1}(\alpha/2 - O(\rho_n^{-1} n^{-1})) \leqslant \widehat{T}_n \leqslant G_n^{-1}(\alpha/2 + O(\rho_n^{-1} n^{-1}))\right) \\
&\qquad\quad + O(n^{-1}) \\
(\text{Theorem 3.1}) &\qquad = G_n(G_n^{-1}(\alpha/2 + O(\rho_n^{-1} n^{-1}))) - G_n(G_n^{-1}(\alpha/2 - O(\rho_n^{-1} n^{-1}))) \\
&\qquad\quad + O\left(\mathcal{M}(\rho_n, n; R)\right) = O\left(\mathcal{M}(\rho_n, n; R)\right).
\end{aligned}$$

Now we continue (8.122) and bound $\mathbb{P}\left(G_n(\widehat{T}_n) \leqslant \alpha/2\right)$. We have

$$\begin{aligned}
\mathbb{P}\Big(G_n(\widehat{T}_n) \leqslant \alpha/2\Big) =& \mathbb{P}\left(G_n(\widehat{T}_n) \leqslant \alpha/2, \widehat{T}_n \in [z_{\alpha/2} - \delta_0, z_{\alpha/2} + \delta_0]\right) \\
& + \mathbb{P}\left(G_n(\widehat{T}_n) \leqslant \alpha/2, \widehat{T}_n > z_{\alpha/2} + \delta_0\right) \\
& + \mathbb{P}\left(G_n(\widehat{T}_n) \leqslant \alpha/2, \widehat{T}_n < z_{\alpha/2} - \delta_0\right) \\
(\text{Property (i)}) =& \mathbb{P}\left(G_n(\widehat{T}_n) \leqslant \alpha/2, \widehat{T}_n \in [z_{\alpha/2} - \delta_0, z_{\alpha/2} + \delta_0]\right) + \mathbb{P}\left(\widehat{T}_n < z_{\alpha/2} - \delta_0\right)
\end{aligned}$$

$$=\mathbb{P}\left(\widehat{T}_n \leqslant G_n^{-1}(\alpha/2), \widehat{T}_n \in [z_{\alpha/2}-\delta_0, z_{\alpha/2}+\delta_0]\right) + \mathbb{P}\left(\widehat{T}_n < z_{\alpha/2}-\delta_0\right)$$

$$(\text{Since } G_n^{-1}(\alpha/2) \geqslant z_{\alpha/2}-\delta_0) = \mathbb{P}\left(\widehat{T}_n \leqslant G_n^{-1}(\alpha/2)\right) = F_{\widehat{T}_n}(G_n^{-1}(\alpha/2))$$

$$=G_n(G_n^{-1}(\alpha/2)) + O\left(\mathcal{M}(\rho_n, n; R)\right) = \alpha/2 + O\left(\mathcal{M}(\rho_n, n; R)\right)$$

The other term $\mathbb{P}\left(G_n(\widehat{T}_n) \geqslant 1-\alpha/2\right)$ can be handled exactly similarly, and the proof of part 1 of Theorem 4.2 is completed.

Now we move on to prove part 2 of the theorem. $|c_n - d_n| = \omega(\rho_n^s \cdot n^{-1/2})$. When H_a is true, we have $\mu_n = d_n$, and rewrite

$$\widehat{T}_n := \frac{\widehat{U}_n - d_n}{\widehat{S}_n} + \frac{d_n - c_n}{\widehat{S}_n}$$

Since $\widehat{S}_n = \widetilde{O}_p(\rho_n^s \cdot n^{-1/2})$, we have

$$\left|\frac{d_n - c_n}{\widehat{S}_n}\right| \xrightarrow{p} \infty, \qquad \text{and therefore,} \qquad |\widehat{T}_n| \xrightarrow{p} \infty$$

By definition of Type-II error, this finishes the proof of part 2 of Theorem 4.2.

□

PROOF OF THEOREM 4.3. We first prove (4.6). By definition, we have

$$\left|F_{\widehat{T}_n}(q_{\widehat{T}_n;\alpha}) - \alpha\right| = F_{\widehat{T}_n}(q_{\widehat{T}_n;\alpha}) - \alpha \leqslant F_{\widehat{T}_n}(q_{\widehat{T}_n;\alpha}) - F_{\widehat{T}_n}(q_{\widehat{T}_n;\alpha} - 0^+)$$

$$\leqslant \left|F_{\widehat{T}_n}(q_{\widehat{T}_n;\alpha}) - G_n(q_{\widehat{T}_n;\alpha})\right| + \left|G_n(q_{\widehat{T}_n;\alpha}) - G_n(q_{\widehat{T}_n;\alpha} - 0^+)\right|$$

$$+ \left|G_n(q_{\widehat{T}_n;\alpha} - 0^+) - F_{\widehat{T}_n}(q_{\widehat{T}_n;\alpha} - 0^+)\right|$$

$$\leqslant O\left(\mathcal{M}(\rho_n, n; R)\right) + 0^+ = O\left(\mathcal{M}(\rho_n, n; R)\right)$$

where 0^+ represents an arbitrarily small positive number that may depend on n, and in the last line we used the fact that $G_n(x)$ is globally uniformly Lipschitz. This proves (4.6).

Then we prove the horizontal bound (4.7). Define

$$\widetilde{q}_{\widehat{T}_n;\alpha} := z_\alpha - \frac{1}{\sqrt{n}\cdot \xi_1^3} \cdot \left\{ \frac{2z_\alpha^2+1}{6} \cdot \mathbb{E}[g_1^3(X_1)] \right.$$

$$\left. + \frac{r-1}{2} \cdot \left(z_\alpha^2+1\right) \mathbb{E}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)] \right\}$$

For convenience, let us simply denote the $n^{-1/2}$ term in the Edgeworth expansion by $\Gamma(x)$:

$$\Gamma(x) := \frac{1}{\xi_1^3} \cdot \left\{ \frac{2x^2+1}{6} \cdot \mathbb{E}[g_1^3(X_1)] + \frac{r-1}{2} \cdot \left(x^2+1\right) \mathbb{E}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)] \right\}$$

$$\widehat{\Gamma}(x) := \frac{1}{\widehat{\xi}_1^3} \cdot \left\{ \frac{2x^2+1}{6} \cdot \widehat{\mathbb{E}}[g_1^3(X_1)] + \frac{r-1}{2} \cdot \left(x^2+1\right) \widehat{\mathbb{E}}[g_1(X_1)g_1(X_2)g_2(X_1,X_2)] \right\}$$

We have

$$G_n(x) = \Phi(x) + n^{-1/2} \cdot \Gamma(x)\varphi(x)$$

$$\widetilde{q}_{\widehat{T}_n;\alpha} = z_\alpha - n^{-1/2} \cdot \Gamma(z_\alpha)$$

$$\widehat{q}_{\widehat{T}_n;\alpha} = z_\alpha - n^{-1/2} \cdot \widehat{\Gamma}(z_\alpha)$$

Then the proof of Theorem 3.2 immediately implies that $\left|\widehat{q}_{\widehat{T}_n;\alpha} - \widetilde{q}_{\widehat{T}_n;\alpha}\right| = \widetilde{O}_p(\rho_n^{-1} n^{-1} \log^{1/2} n)$. Mimicking the inversion formula in [65], we have

$$G_n\left(x - \frac{1}{\sqrt{n}} \cdot \Gamma(x)\right) = \Phi\left(x - \frac{1}{\sqrt{n}} \cdot \Gamma(x)\right) + \frac{1}{\sqrt{n}} \cdot \Gamma\left(x - \frac{1}{\sqrt{n}} \cdot \Gamma(x)\right) \varphi\left(x - \frac{1}{\sqrt{n}} \cdot \Gamma(x)\right)$$

$$= \Phi(x) + O\left(n^{-1}\right). \tag{8.123}$$

Also notice that the remainder bound in (8.123) holds uniformly over all $x \in \mathbb{R}$. As [65] pointed out, in a more general setting, the inversion formula (8.123) might not always have a uniform $O(n^{-1})$ error bound, when the leading term in the Edgeworth expansion contains a jump function component, in which case the uniform error bound of the Cornish-Fisher expansion is just $O(n^{-1/2})$. But in our setting, $\Gamma(x)$ is always continuous, and moreover, Lipscitz, so [65]'s remark would not be a concern.

We continue our proof. By Theorem 3.1 and (8.123), we have

$$G_n(\widetilde{q}_{\widehat{T}_n;\alpha}) = \alpha + O(n^{-1}) \tag{8.124}$$

$$G_n(q_{\widehat{T}_n;\alpha}) = F_{\widehat{T}_n}(q_{\widehat{T}_n;\alpha}) + O\left(\mathcal{M}(\rho_n, n; R)\right) \tag{8.125}$$

Since for any given α and large enough n, properties (i) and (ii) of $G_n(\cdot)$, with "$\alpha/2$, ϵ_0, δ_0, n_0" replaced by "α, ϵ_0', δ_0', n_0'", around a neighborhood of z_α. This yields that for large enough n, both $q_{\widehat{T}_n;\alpha}$ and $\widetilde{q}_{\widehat{T}_n;\alpha}$ belong to $[z_\alpha - \epsilon_0', z_\alpha + \epsilon_0']$. Then using the invertibility and the Lipschitz property of the inverse function of $G_n(\cdot)$ within this compact neighborhood, we have

$$|\widetilde{q}_{\widehat{T}_n;\alpha} - q_{\widehat{T}_n;\alpha}| \lesssim |G_n(\widetilde{q}_{\widehat{T}_n;\alpha}) - G_n(q_{\widehat{T}_n;\alpha})|$$

$$= O\left(\mathcal{M}(\rho_n, n; R)\right) \tag{8.126}$$

Combining this with the error bound on $|\widetilde{q}_{\widehat{T}_n;\alpha} - \widehat{q}_{\widehat{T}_n;\alpha}|$ we obtained earlier finishes the proof of the horizontal error bound (4.7).

Now we prove the vertical error bound (4.8). Here we should be careful that $\mathbb{P}(\widehat{T}_n \leqslant \widehat{q}_{\widehat{T}_n;\alpha})$ does *not* equal $F_{\widehat{T}_n}(\widehat{q}_{\widehat{T}_n;\alpha})$, as the former is non-random and the latter is random. In order to study $\mathbb{P}(\widehat{T}_n \leqslant \widehat{q}_{\widehat{T}_n;\alpha})$, we seek the help from $\widetilde{q}_{\widehat{T}_n;\alpha}$ and appeal to the basic definition. By the horizontal error bound, we know that with probability $1 - O(n^{-1})$, we have

$$|\widetilde{q}_{\widehat{T}_n;\alpha} - \widehat{q}_{\widehat{T}_n;\alpha}| \leqslant C \cdot \mathcal{M}(\rho_n, n; R)$$

for some constant $C > 0$. This yields that under the above event

$$F_{\widehat{T}_n}(\widetilde{q}_{\widehat{T}_n;\alpha} - C \cdot \mathcal{M}(\rho_n, n; R)) = \mathbb{P}(\widehat{T}_n \leqslant \widetilde{q}_{\widehat{T}_n;\alpha} - C \cdot \mathcal{M}(\rho_n, n; R))$$

$$\leqslant \mathbb{P}(\widehat{T}_n \leqslant \widehat{q}_{\widehat{T}_n;\alpha})$$

$$\leqslant \mathbb{P}(\widehat{T}_n \leqslant \widetilde{q}_{\widehat{T}_n;\alpha} + C \cdot \mathcal{M}(\rho_n, n; R)) = F_{\widehat{T}_n}(\widetilde{q}_{\widehat{T}_n;\alpha} + C \cdot \mathcal{M}(\rho_n, n; R))$$

Recall that $G_n(\cdot)$ is globally Lipscitz for large enough n, we have

$$F_{\widehat{T}_n}(\widetilde{q}_{\widehat{T}_n;\alpha} + C \cdot \mathcal{M}(\rho_n, n; R)) = G_n(\widetilde{q}_{\widehat{T}_n;\alpha} + C \cdot \mathcal{M}(\rho_n, n; R)) + O\left(\mathcal{M}(\rho_n, n; R)\right)$$

$$= G_n(\widetilde{q}_{\widehat{T}_n;\alpha}) + O\left(\mathcal{M}(\rho_n, n; R)\right) = \alpha + O\left(\mathcal{M}(\rho_n, n; R)\right)$$

This proves the vertical error bound (4.8) and concludes the proof of Theorem 4.3.

□

9. Additional simulation results.

9.1. *Additional results in Simulation 5.1.* In this section, we show additional simulation results under different network sparsity settings. We tested $\rho_n \asymp n^{-1/4}, n^{-1/3}$ and $n^{-1/2}$. Notice that some of these settings constitute violations of our assumptions ρ_n assumptions. We adjusted the constant factors in ρ_n such that all settings start with roughly equal network densities for $n = 10$. Results are shown in Figures 6–8 (errors) and Figures 9–11 (time costs), where error bars show standard deviations.

The plots show that the accuracy of all methods depreciate as the network becomes sparser. Recall that our loss function is the error in approximating $F_{\widehat{T}_n}$, and that $\widehat{T}_n$ is normalized by the denominator $\widehat{S}_n \asymp \rho_n^s \cdot n^{-1/2}$, it is therefore understandable that sparser networks are more difficult. Apart from that error bounds would depreciate with a smaller ρ_n, as in our Theorems 3.1 and 3.2; the performances of our method in some scenarios also seemed to be limited by numerical accuracy, possibly in the Monte Carlo evaluations of the true $F_{\widehat{T}_n}$. But overall, our method remains the best performer and higher-order accurate in scenarios where the sparsity assumptions are satisfied. The time cost plots can be interpreted similarly to that in the main paper text.

9.2. *Additional results in Simulation 5.2.* In this subsection we present the results for more settings, including $n = 160$ and more sparsity levels. Results are reported in Tables 6–23.

TABLE 6
Performance measures of 95% confidence intervals
$n = 80$, $\rho_n \asymp n^{-1/4}$, graphon: block model

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.960(0.196)$	$0.954(0.209)$	$0.957(0.203)$	$0.953(0.212)$
	Length $= 0.084(0.009)$	$0.024(0.005)$	$0.144(0.024)$	$0.087(0.020)$
	LogTime $= -8.419(0.135)$	$-7.450(0.118)$	$-7.404(0.108)$	$-6.405(0.774)$
Norm. Approx.	$0.953(0.212)$	$0.935(0.247)$	$0.944(0.230)$	$0.933(0.251)$
	$0.084(0.009)$	$0.024(0.005)$	$0.144(0.024)$	$0.087(0.020)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.830(0.376)$	$0.856(0.351)$	$0.832(0.374)$	$0.858(0.349)$
	$0.058(0.008)$	$0.019(0.004)$	$0.106(0.019)$	$0.069(0.016)$
	$-2.599(0.028)$	$-2.137(0.020)$	$-2.195(0.031)$	$-0.987(0.015)$
Green and Shalizi [61]	$0.934(0.249)$	$0.936(0.245)$	$0.942(0.234)$	$0.938(0.241)$
	$0.082(0.011)$	$0.027(0.006)$	$0.145(0.028)$	$0.089(0.023)$
	$-1.202(0.019)$	$0.548(0.051)$	$0.085(0.052)$	$0.353(0.012)$
Levin and Levina [93]	$0.954(0.210)$	$0.956(0.205)$	$0.956(0.205)$	$0.952(0.214)$
	$0.085(0.011)$	$0.026(0.006)$	$0.150(0.028)$	$0.094(0.023)$
	$-1.193(0.014)$	$0.574(0.040)$	$0.074(0.044)$	$0.403(0.006)$

9.3. *Additional results in Simulation 5.3.* In this simulation, all settings are carried over from Simulation 5.3 except that $n = 80$. The results are shown in Figure 12. We observed the anticipated depreciation in the performances of all methods, while our method maintains a consistent advantage over the closest competitors.

The impact of ρ_n on the computation time is a subtle topic. Since our simulation runs across dense and sparse regimes, for simplicity and wide-applicability of the code, we did not engage sparse matrix computation procedures. Consequently, the time cost for all ρ_n's are nearly the same for all methods. Here, we only show the time costs for $n = 80$ in Figure 13, and the analogous plot for $n = 160$ looks exactly similar and is thus omitted here. We leave the study of improving computational efficiency for sparse ρ_n's to future work.

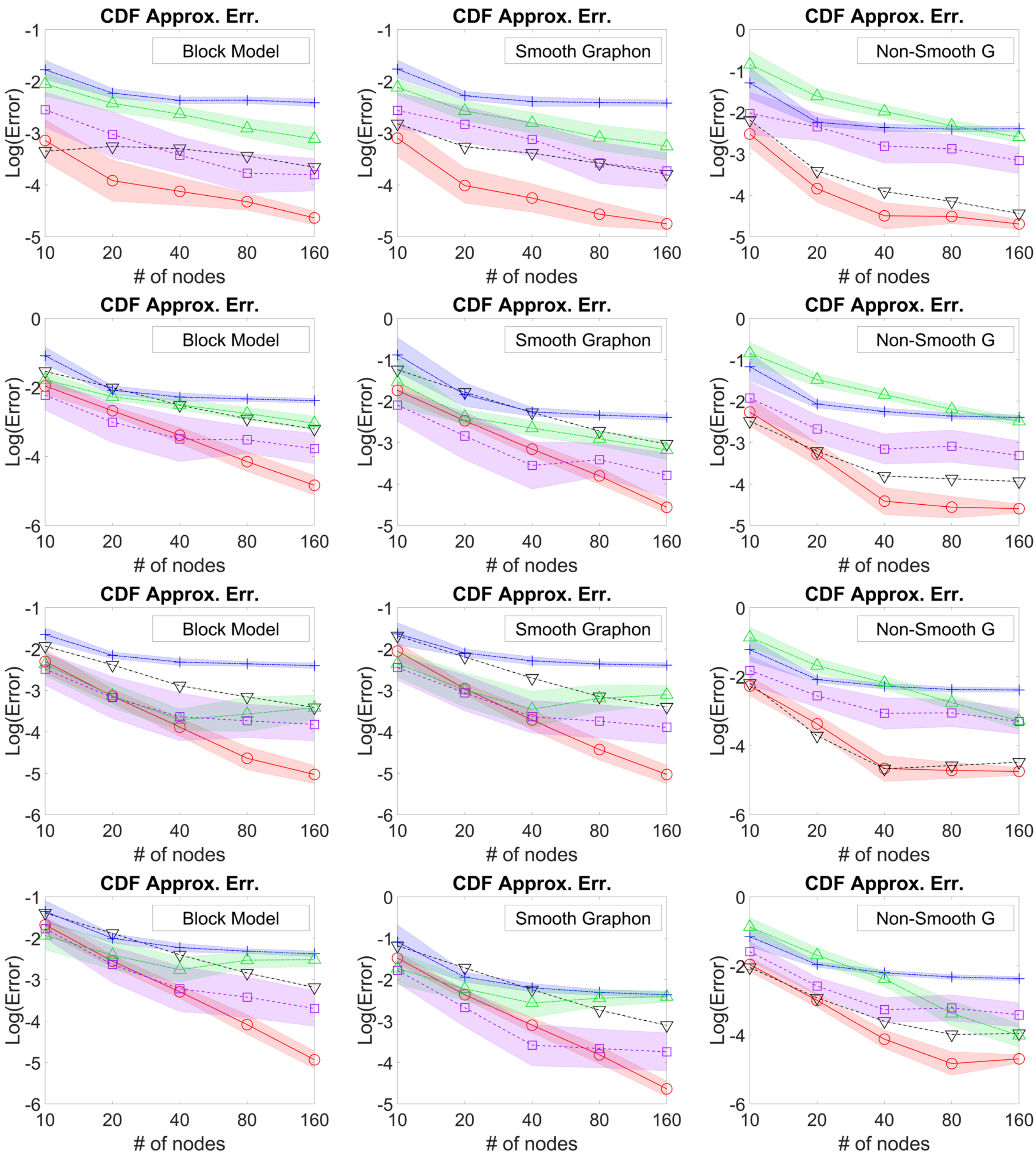

Fig 6: CDF approximation errors, $\rho_n \asymp n^{-1/4}$. Both axes are log(e)-scaled. **Motifs:** row 1: `Edge`; row 2: `Triangle`; row 3: `Vshape`; row 4: `ThreeStar`. Red solid curve marked circle: our method (empirical Edgeworth); black dashed curve marked down-triangle: $N(0,1)$ approximation; green dashed curve marked up-triangle: re-sampling of A in [61]; blue dashed curve marked plus: [17] sub-sampling $\asymp n$ nodes; magenta dashed line with square markers: ASE plug-in bootstrap in [93].

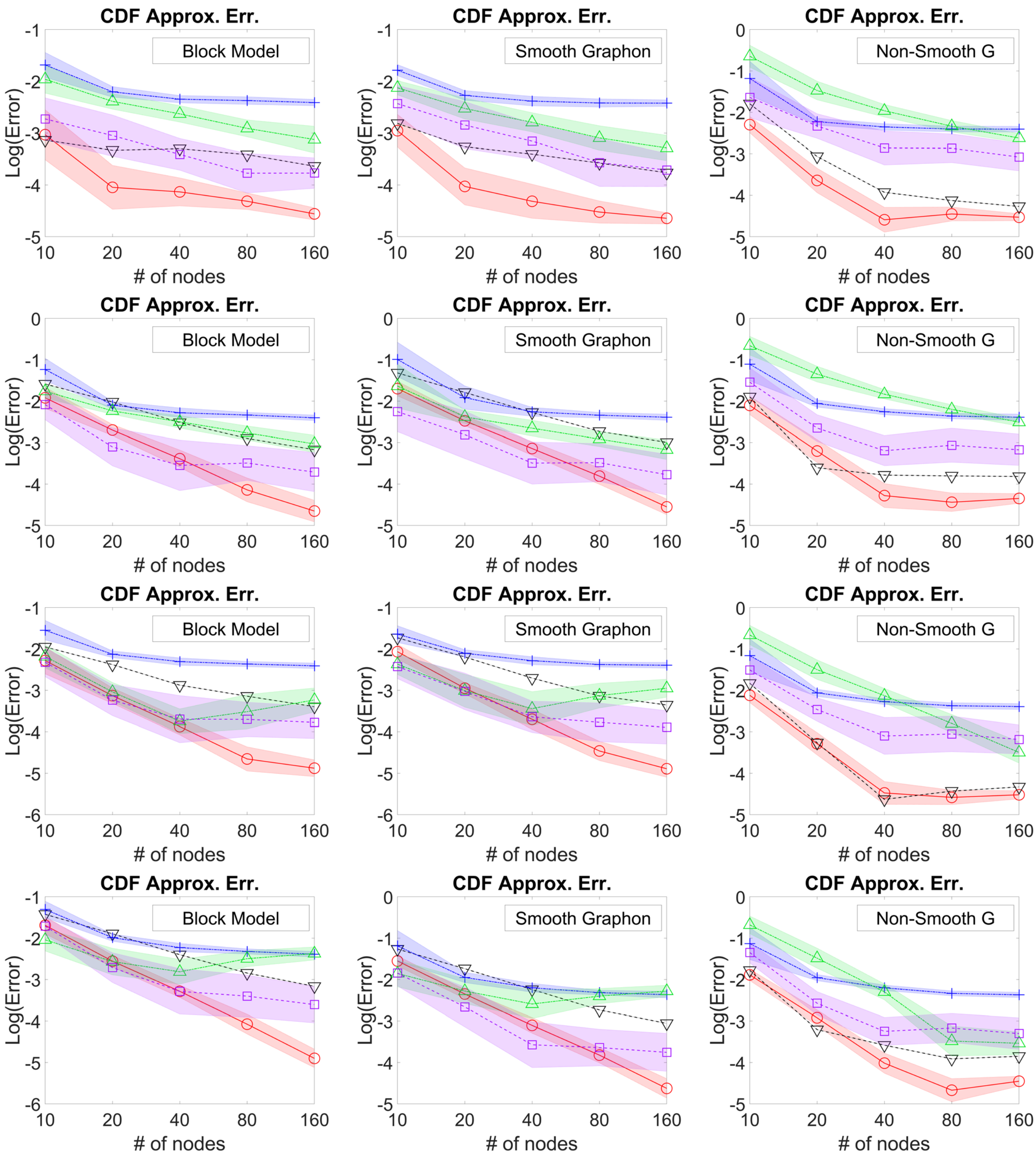

Fig 7: CDF approximation errors, $\rho_n \asymp n^{-1/3}$. Both axes are log(e)-scaled. **Motifs:** row 1: `Edge`; row 2: `Triangle`; row 3: `Vshape`; row 4: `ThreeStar`. Red solid curve marked circle: our method (empirical Edgeworth); black dashed curve marked down-triangle: $N(0,1)$ approximation; green dashed curve marked up-triangle: re-sampling of A in [61]; blue dashed curve marked plus: [17] sub-sampling $\asymp n$ nodes; magenta dashed line with square markers: ASE plug-in bootstrap in [93].

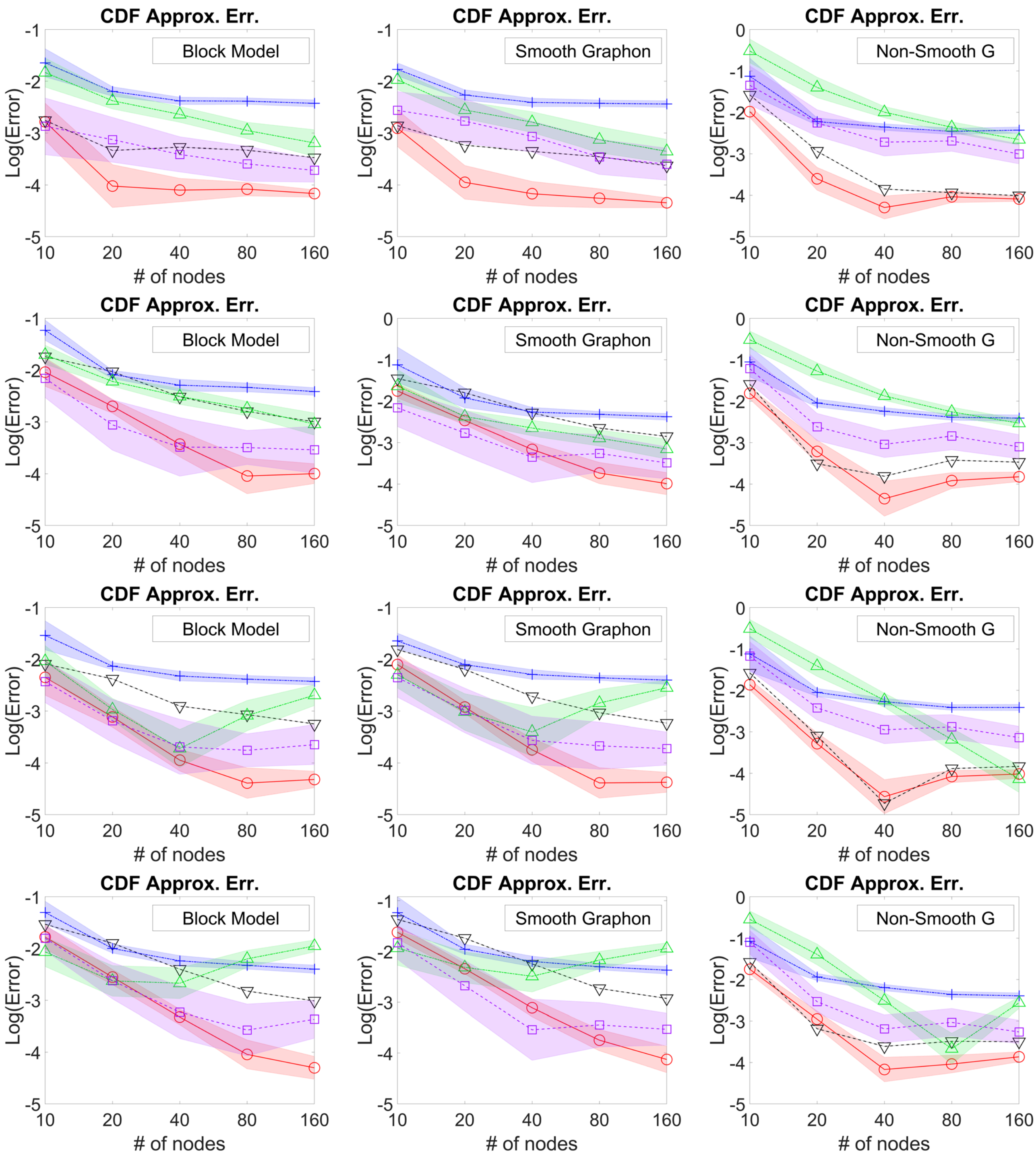

Fig 8: CDF approximation errors, $\rho_n \asymp n^{-1/2}$. Both axes are log(e)-scaled. **Motifs:** row 1: `Edge`; row 2: `Triangle`; row 3: `Vshape`; row 4: `ThreeStar`. Red solid curve marked circle: our method (empirical Edgeworth); black dashed curve marked down-triangle: $N(0,1)$ approximation; green dashed curve marked up-triangle: re-sampling of A in [61]; blue dashed curve marked plus: [17] sub-sampling $\asymp n$ nodes; magenta dashed line with square markers: ASE plug-in bootstrap in [93].

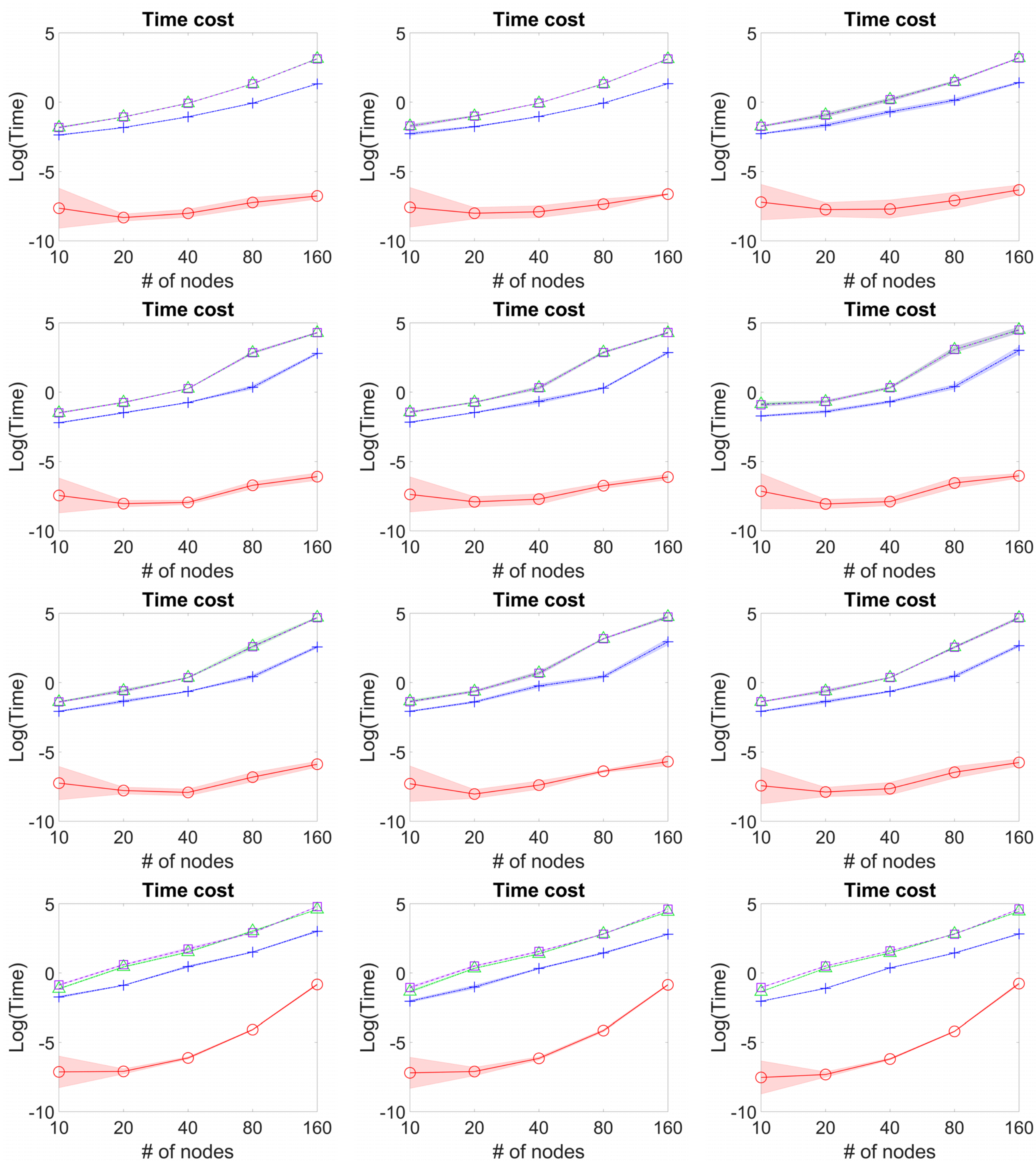

Fig 9: CDF approximation times, $\rho_n \asymp n^{-1/4}$. Both axes are log(e)-scaled. **Motifs:** row 1: `Edge`; row 2: `Triangle`; row 3: `Vshape`; row 4: `ThreeStar`. Red solid curve marked circle: our method (empirical Edgeworth); black dashed curve marked down-triangle: $N(0,1)$ approximation; green dashed curve marked up-triangle: re-sampling of A in [61]; blue dashed curve marked plus: [17] sub-sampling $\asymp n$ nodes; magenta dashed line with square markers: ASE plug-in bootstrap in [93].

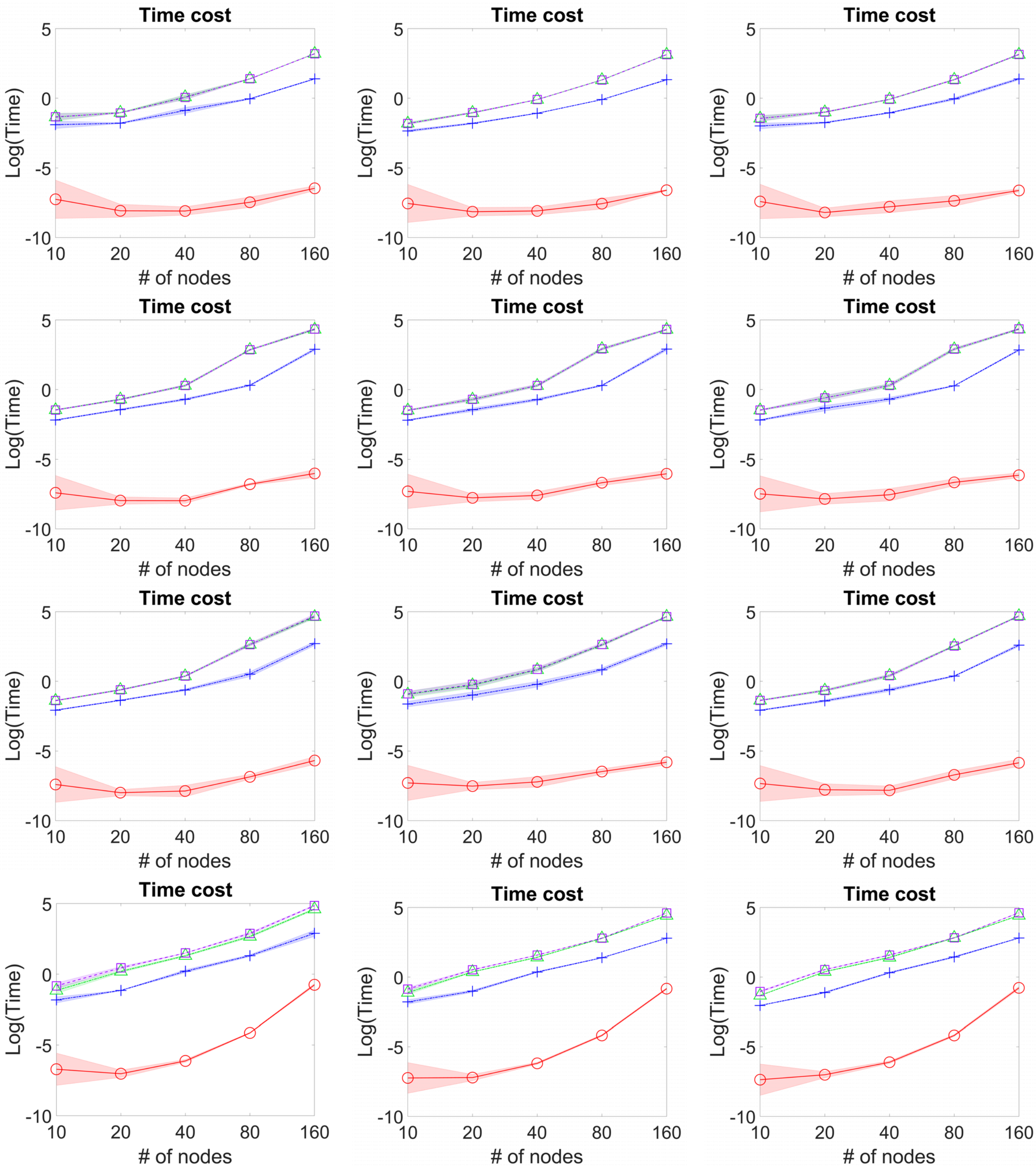

Fig 10: CDF approximation times, $\rho_n \asymp n^{-1/3}$. Both axes are log(e)-scaled. **Motifs:** row 1: `Edge`; row 2: `Triangle`; row 3: `Vshape`; row 4: `ThreeStar`. Red solid curve marked circle: our method (empirical Edgeworth); black dashed curve marked down-triangle: $N(0,1)$ approximation; green dashed curve marked up-triangle: re-sampling of A in [61]; blue dashed curve marked plus: [17] sub-sampling $\asymp n$ nodes; magenta dashed line with square markers: ASE plug-in bootstrap in [93].

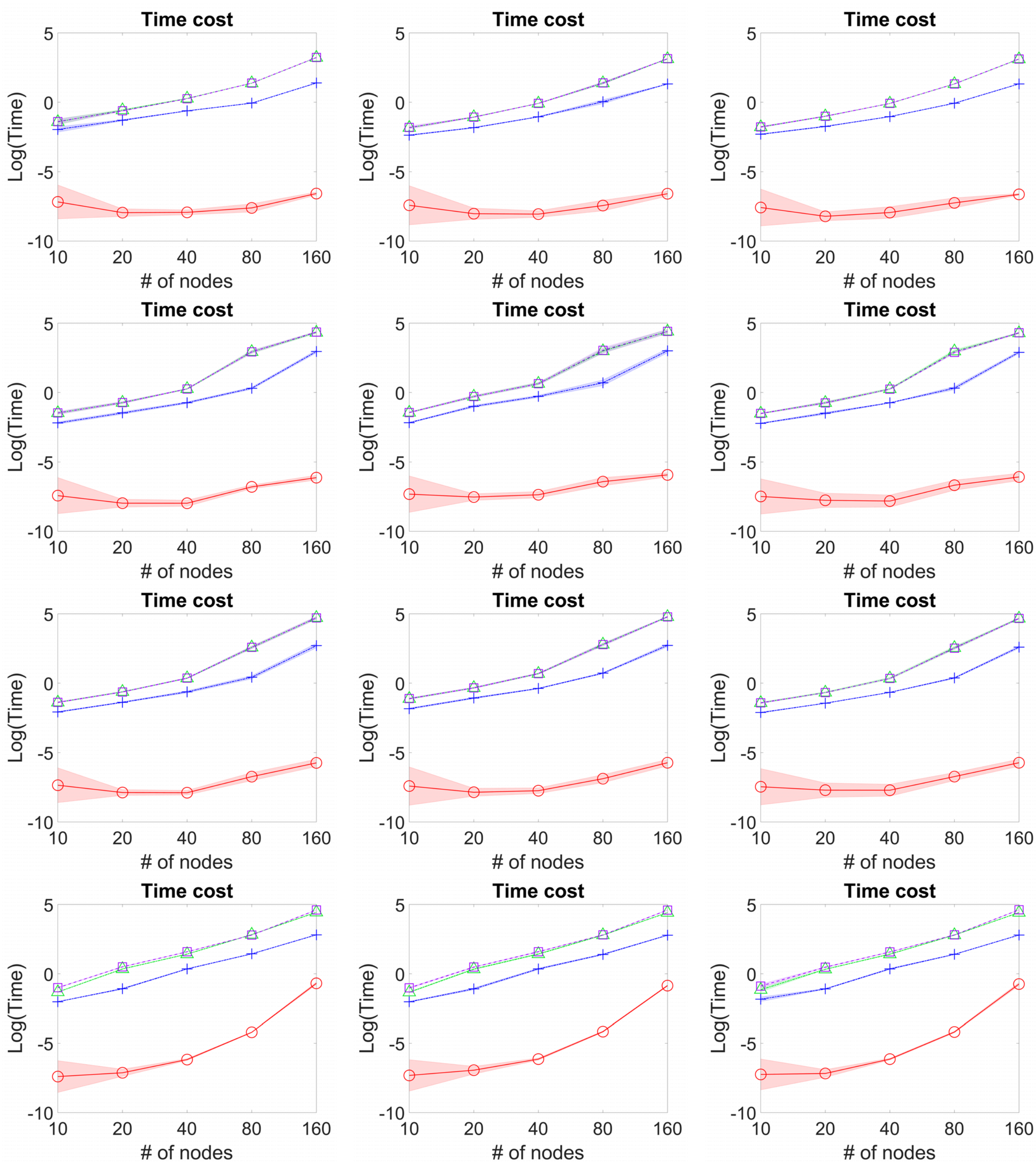

Fig 11: CDF approximation times, $\rho_n \asymp n^{-1/2}$. Both axes are log(e)-scaled. **Motifs:** row 1: `Edge`; row 2: `Triangle`; row 3: `Vshape`; row 4: `ThreeStar`. Red solid curve marked circle: our method (empirical Edgeworth); black dashed curve marked down-triangle: $N(0,1)$ approximation; green dashed curve marked up-triangle: re-sampling of A in [61]; blue dashed curve marked plus: [17] sub-sampling $\asymp n$ nodes; magenta dashed line with square markers: ASE plug-in bootstrap in [93].

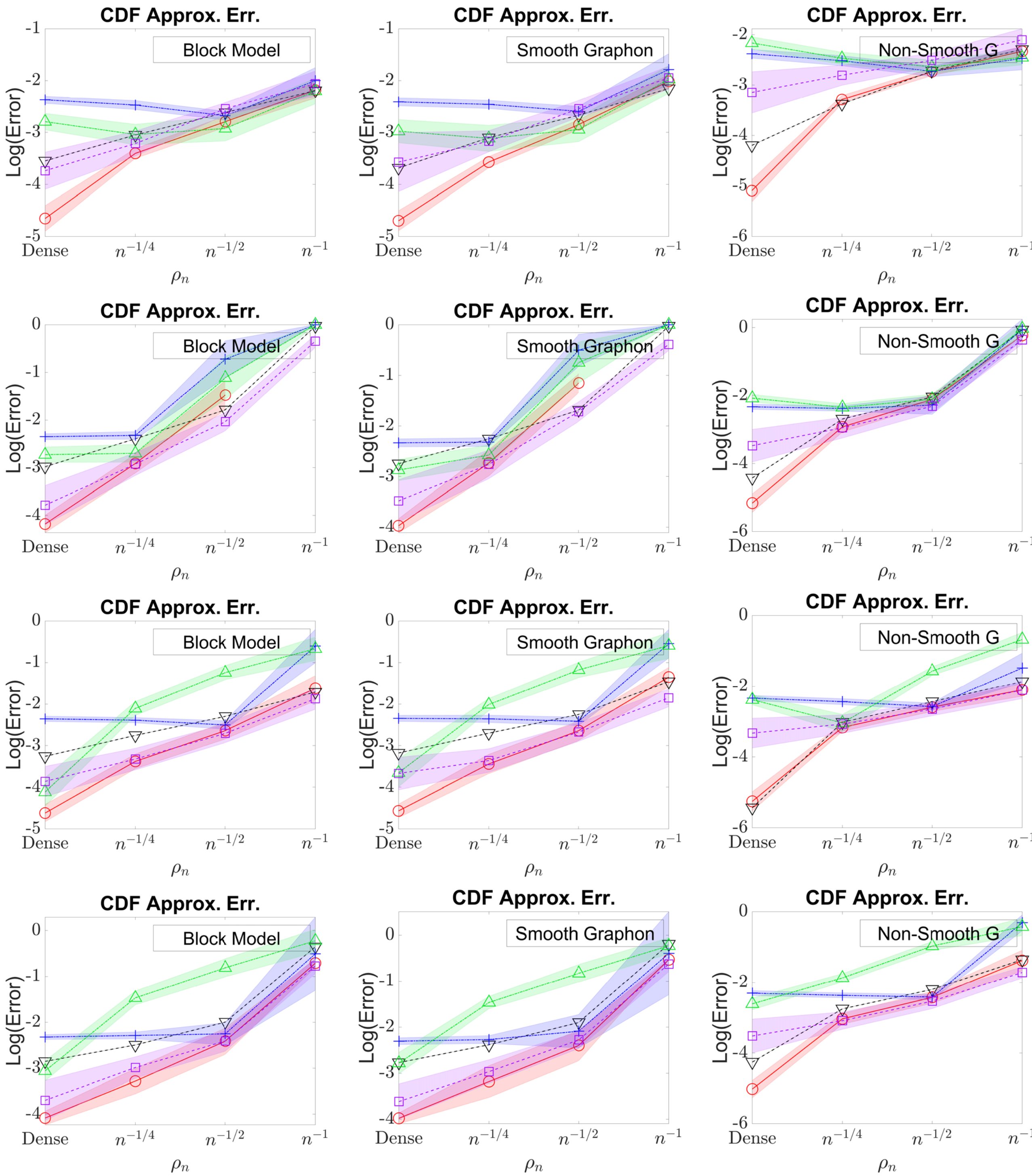

Fig 12: Impact of sparsity on approximation errors, $n = 80$. Both axes are log(e)-scaled. **Motifs:** row 1: `Edge`; row 2: `Triangle`; row 3: `Vshape`; row 4: `ThreeStar`. Red solid curve marked circle: our method (empirical Edgeworth); black dashed curve marked down-triangle: $N(0,1)$ approximation; green dashed curve marked up-triangle: re-sampling of A in [61]; blue dashed curve marked plus: [17] sub-sampling $\asymp n$ nodes; magenta dashed line with square markers: ASE plug-in bootstrap in [93]. We regarded $N(0,1)$ as zero time cost so does not appear in the time cost plot.

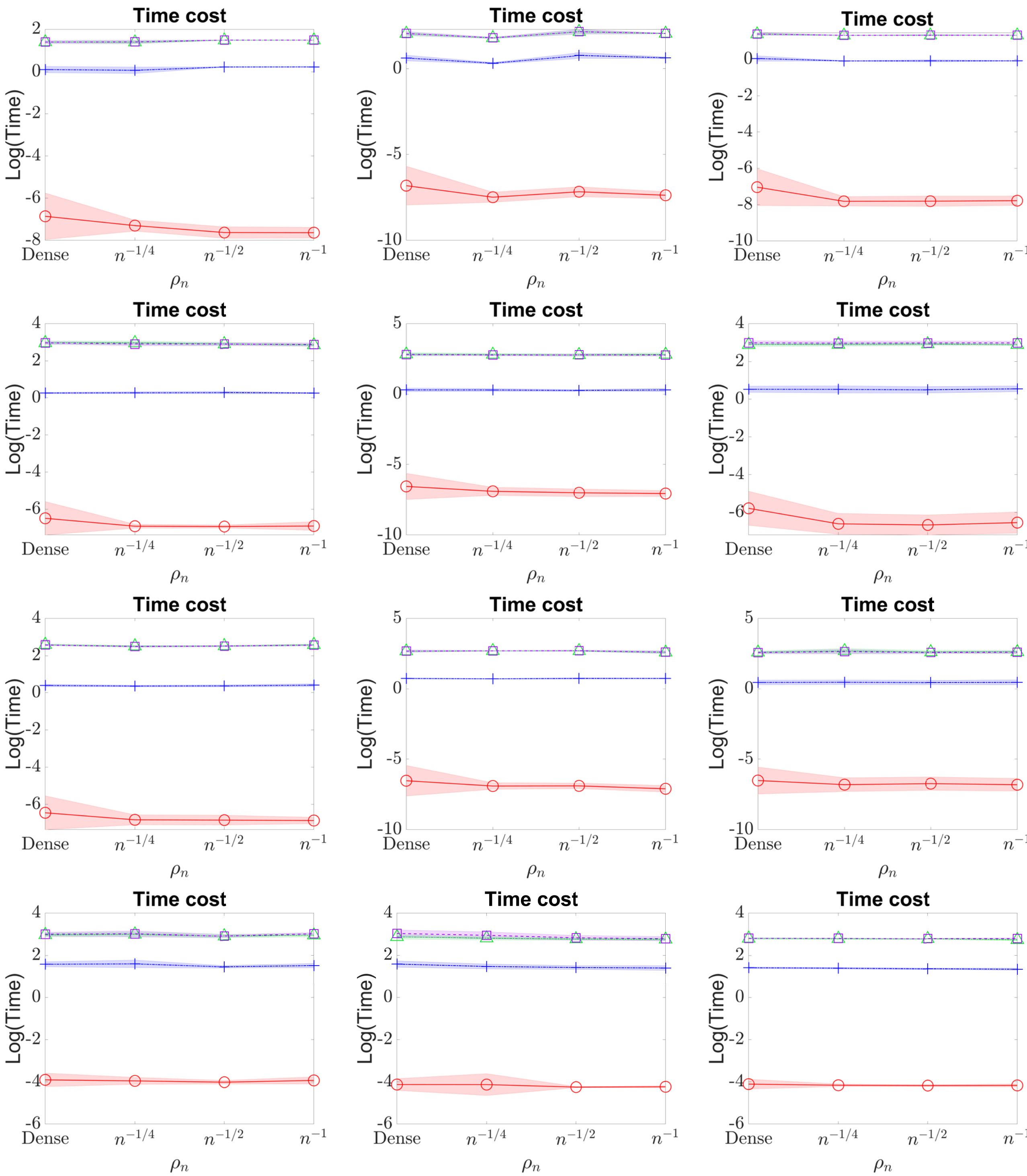

Fig 13: Impact of sparsity on time cost, $n = 80$. We used regular (non-sparse) matrix variables in MATLAB. Both axes are log(e)-scaled. **Motifs:** row 1: `Edge`; row 2: `Triangle`; row 3: `Vshape`; row 4: `ThreeStar`. Red solid curve marked circle: our method (empirical Edgeworth); green dashed curve marked up-triangle: re-sampling of A in [61]; blue dashed curve marked plus: [17] sub-sampling $\asymp n$ nodes; magenta dashed line with square markers: ASE plug-in bootstrap in [93]. We regarded $N(0, 1)$ as zero time cost so does not appear in the time cost plot.

TABLE 7
Performance measures of 95% confidence intervals
$n = 80$, $\rho_n \asymp n^{-1/4}$, *graphon: smooth graphon*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.961(0.193)$	$0.942(0.234)$	$0.955(0.208)$	$0.945(0.229)$
	Length $= 0.078(0.008)$	$0.013(0.003)$	$0.101(0.018)$	$0.050(0.013)$
	LogTime $= -8.226(0.044)$	$-7.468(0.116)$	$-7.439(0.073)$	$-6.599(0.687)$
Norm. Approx.	$0.953(0.211)$	$0.922(0.268)$	$0.939(0.239)$	$0.923(0.266)$
	$0.078(0.008)$	$0.013(0.003)$	$0.101(0.018)$	$0.050(0.013)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.822(0.383)$	$0.850(0.357)$	$0.852(0.355)$	$0.848(0.359)$
	$0.056(0.007)$	$0.011(0.003)$	$0.079(0.015)$	$0.045(0.012)$
	$-2.562(0.008)$	$-2.111(0.099)$	$-2.198(0.044)$	$-0.995(0.012)$
Green and Shalizi [61]	$0.928(0.259)$	$0.948(0.222)$	$0.934(0.249)$	$0.944(0.230)$
	$0.078(0.010)$	$0.015(0.004)$	$0.105(0.021)$	$0.054(0.015)$
	$-1.148(0.010)$	$0.504(0.057)$	$0.104(0.102)$	$0.322(0.017)$
Levin and Levina [93]	$0.942(0.234)$	$0.960(0.196)$	$0.954(0.210)$	$0.962(0.191)$
	$0.082(0.010)$	$0.015(0.004)$	$0.111(0.022)$	$0.058(0.016)$
	$-1.146(0.004)$	$0.514(0.048)$	$0.056(0.055)$	$0.387(0.011)$

TABLE 8
Performance measures of 95% confidence intervals
$n = 80$, $\rho_n \asymp n^{-1/4}$, *graphon: non-smooth graphon*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.960(0.196)$	$0.961(0.193)$	$0.961(0.193)$	$0.963(0.189)$
	Length $= 0.101(0.008)$	$0.083(0.007)$	$0.310(0.022)$	$0.329(0.029)$
	LogTime $= -7.996(0.058)$	$-7.652(0.145)$	$-7.611(0.133)$	$-6.789(0.618)$
Norm. Approx.	$0.957(0.202)$	$0.955(0.208)$	$0.957(0.204)$	$0.956(0.206)$
	$0.101(0.008)$	$0.083(0.007)$	$0.310(0.022)$	$0.329(0.029)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.832(0.374)$	$0.838(0.369)$	$0.834(0.372)$	$0.838(0.369)$
	$0.070(0.008)$	$0.059(0.007)$	$0.216(0.023)$	$0.233(0.027)$
	$-2.559(0.048)$	$-2.151(0.029)$	$-2.129(0.029)$	$-1.000(0.042)$
Green and Shalizi [61]	$0.930(0.255)$	$0.934(0.249)$	$0.944(0.230)$	$0.950(0.218)$
	$0.097(0.011)$	$0.083(0.009)$	$0.301(0.030)$	$0.323(0.036)$
	$-1.152(0.027)$	$0.488(0.054)$	$0.144(0.041)$	$0.341(0.033)$
Levin and Levina [93]	$0.962(0.191)$	$0.972(0.165)$	$0.966(0.181)$	$0.970(0.171)$
	$0.101(0.011)$	$0.086(0.010)$	$0.314(0.031)$	$0.338(0.038)$
	$-1.145(0.027)$	$0.479(0.052)$	$0.141(0.040)$	$0.463(0.023)$

9.4. *Additional simulation results for degree-corrected stochastic block models.* Here we present the simulation results under a degree-corrected stochastic block model [83]. We generate data from the stochastic block model `BlockModel` that we tested in Section 5, with the following degree correction function

$$\theta(x) := |\cos \pi \cdot (x - 1/2)|.$$

The results are reported in Figures 14 – 17. We observe the clear advantage of our method over benchmarks, as predicted by our theory.

REFERENCES

[1] ABBE, E. (2017). Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research* **18** 6446–6531.

[2] AHMED, N. K., NEVILLE, J., ROSSI, R. A. and DUFFIELD, N. (2015). Efficient graphlet counting for large networks. In *2015 IEEE International Conference on Data Mining* 1–10. IEEE.

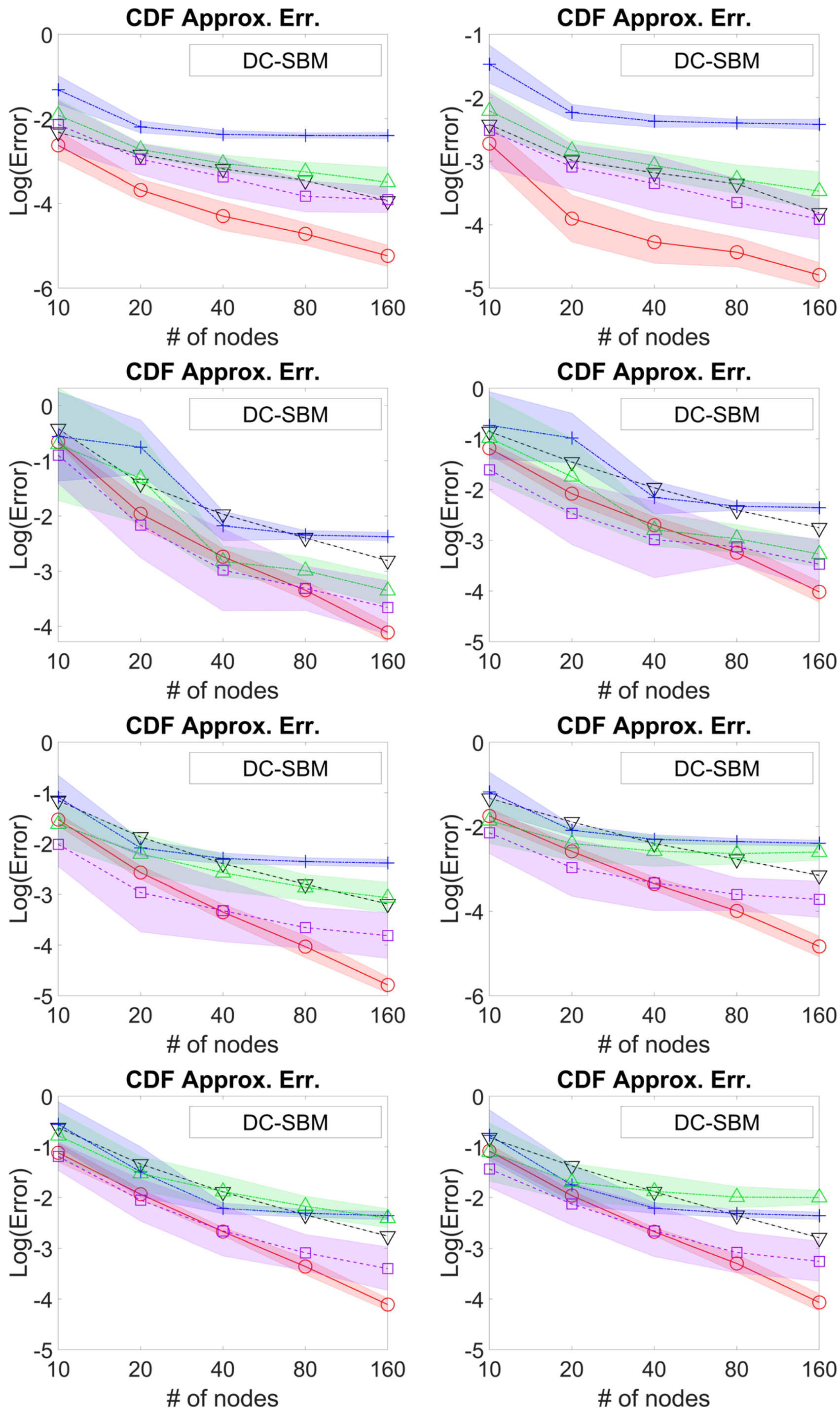

Fig 14: CDF approximation errors for degree-corrected stochastic block model. Sparsity: column 1: $\rho_n \asymp 1$; column 2: $\rho_n \asymp n^{-1/4}$ Both axes are log(e)-scaled. **Motifs:** row 1: `Edge`; row 2: `Triangle`; row 3: `Vshape`; row 4: `ThreeStar`. Red solid curve marked circle: our method (empirical Edgeworth); black dashed curve marked down-triangle: $N(0,1)$ approximation; green dashed curve marked up-triangle: re-sampling of A in [61]; blue dashed curve marked plus: [17] sub-sampling $\asymp n$ nodes; magenta dashed line with square markers: ASE plug-in bootstrap in [93].

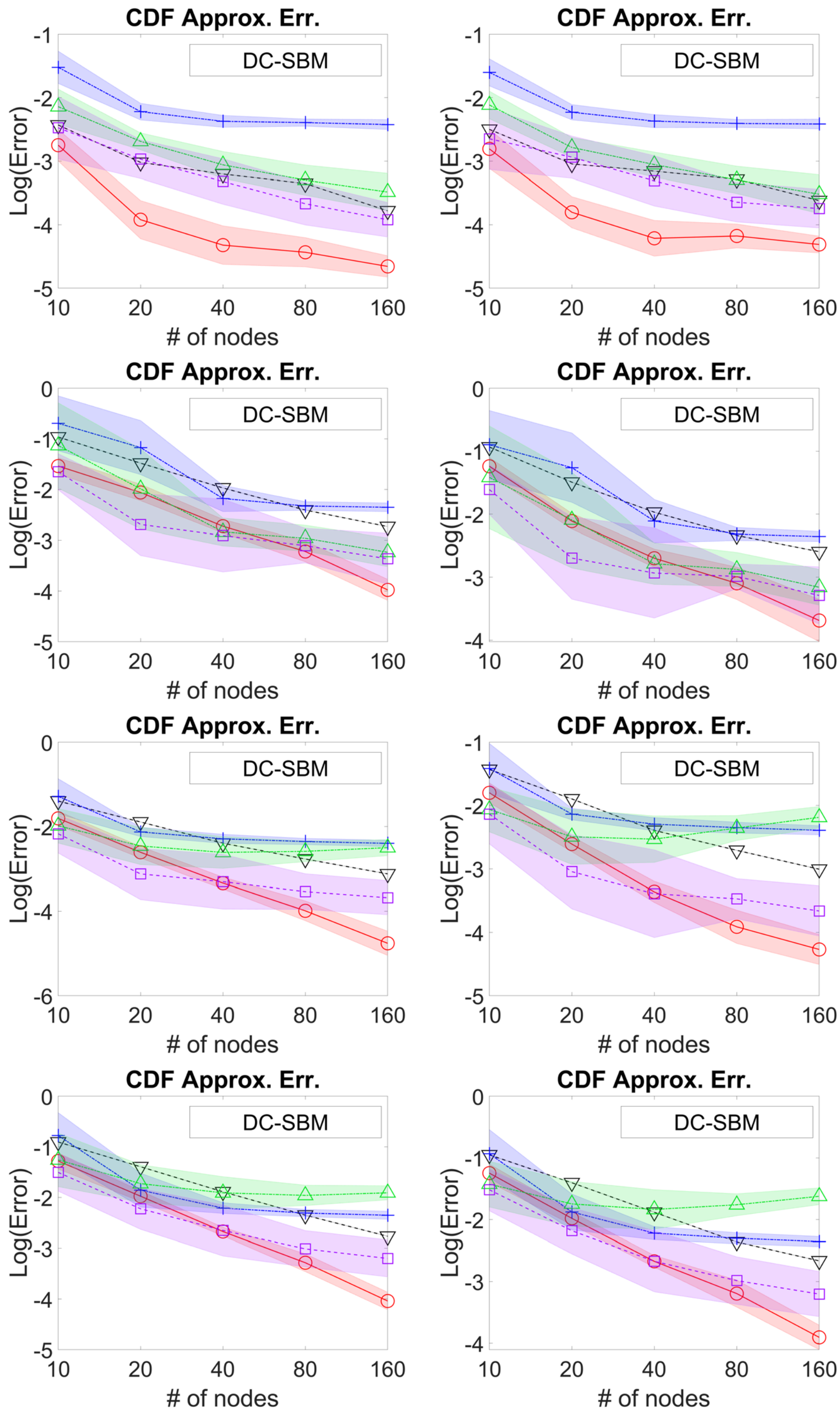

Fig 15: CDF approximation errors for degree-corrected stochastic block model. Sparsity: column 1: $\rho_n \asymp n^{-1/3}$; column 2: $\rho_n \asymp n^{-1/2}$ Both axes are log(e)-scaled. **Motifs:** row 1: `Edge`; row 2: `Triangle`; row 3: `Vshape`; row 4: `ThreeStar`. Red solid curve marked circle: our method (empirical Edgeworth); black dashed curve marked down-triangle: $N(0,1)$ approximation; green dashed curve marked up-triangle: re-sampling of A in [61]; blue dashed curve marked plus: [17] sub-sampling $\asymp n$ nodes; magenta dashed line with square markers: ASE plug-in bootstrap in [93].

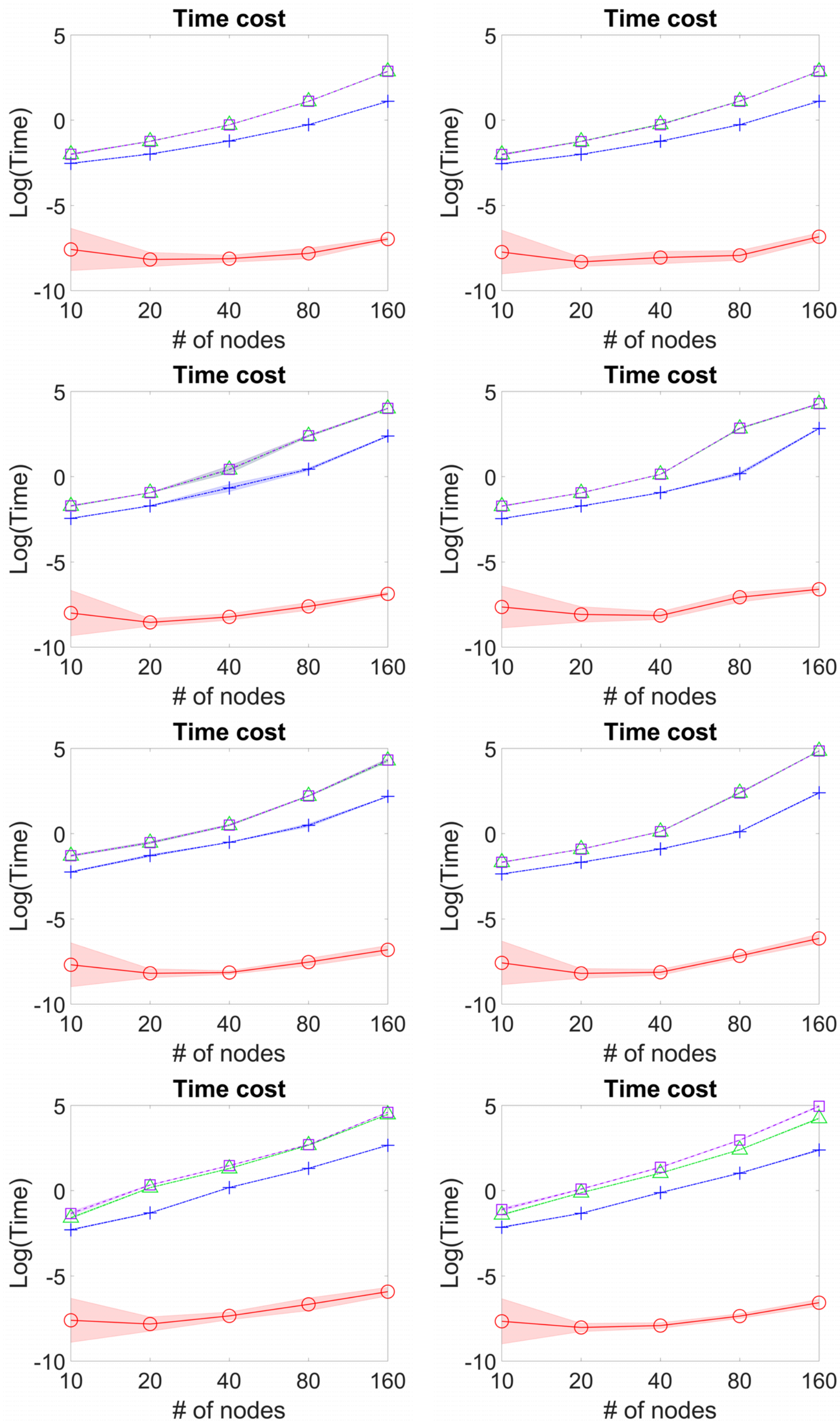

Fig 16: CDF approximation times for degree-corrected stochastic block model. Sparsity: column 1: $\rho_n \asymp 1$; column 2: $\rho_n \asymp n^{-1/4}$ Both axes are log(e)-scaled. **Motifs:** row 1: `Edge`; row 2: `Triangle`; row 3: `Vshape`; row 4: `ThreeStar`. Red solid curve marked circle: our method (empirical Edgeworth); black dashed curve marked down-triangle: $N(0,1)$ approximation; green dashed curve marked up-triangle: re-sampling of A in [61]; blue dashed curve marked plus: [17] sub-sampling $\asymp n$ nodes; magenta dashed line with square markers: ASE plug-in bootstrap in [93].

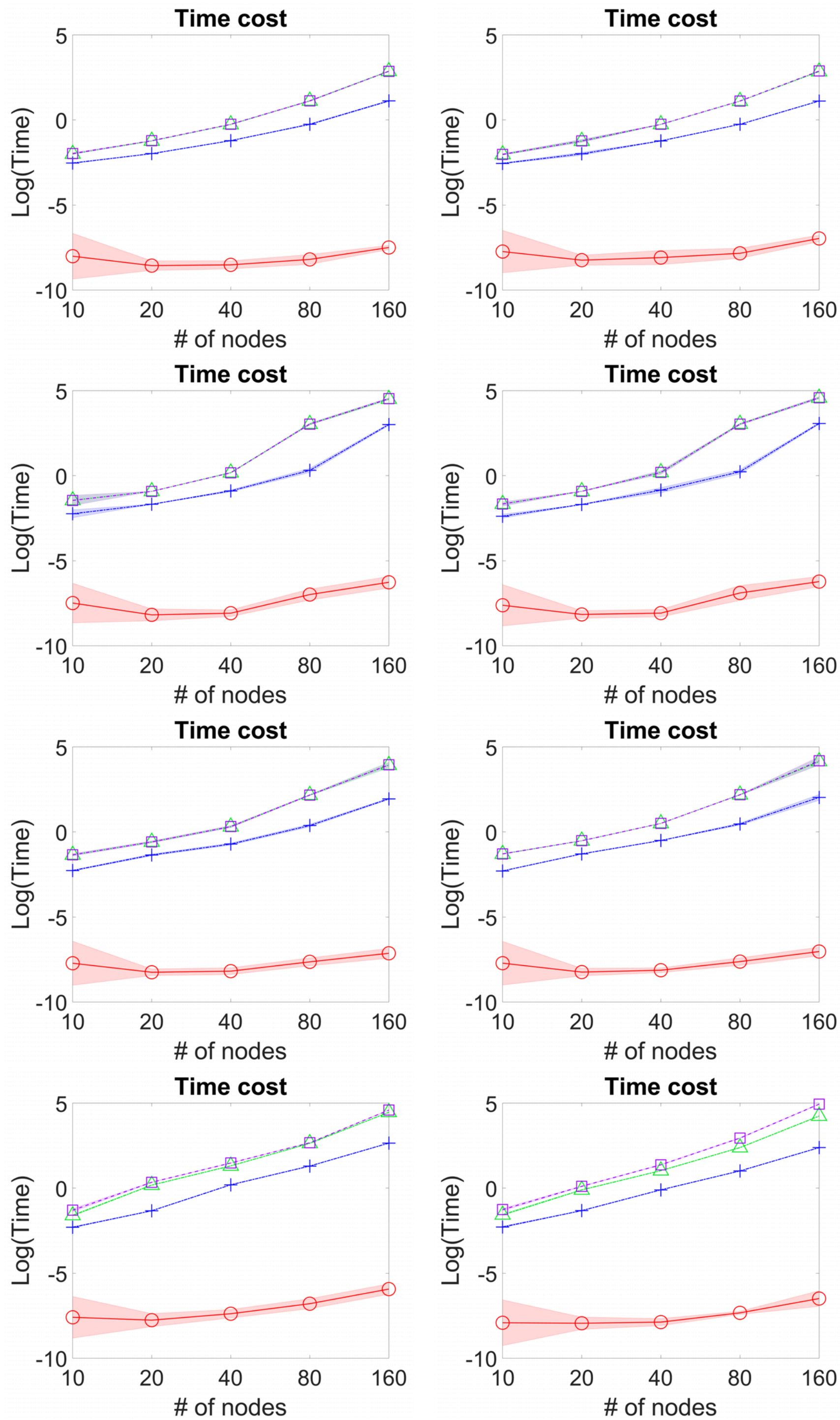

Fig 17: CDF approximation times for degree-corrected stochastic block model. Sparsity: column 1: $\rho_n \asymp n^{-1/3}$; column 2: $\rho_n \asymp n^{-1/2}$ Both axes are log(e)-scaled. **Motifs:** row 1: `Edge`; row 2: `Triangle`; row 3: `Vshape`; row 4: `ThreeStar`. Red solid curve marked circle: our method (empirical Edgeworth); black dashed curve marked down-triangle: $N(0,1)$ approximation; green dashed curve marked up-triangle: re-sampling of A in [61]; blue dashed curve marked plus: [17] sub-sampling $\asymp n$ nodes; magenta dashed line with square markers: ASE plug-in bootstrap in [93].

TABLE 9
Performance measures of 95% confidence intervals
$n = 80$, $\rho_n \asymp n^{-1/2}$, *graphon: block model*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.969(0.172)$	$0.956(0.206)$	$0.964(0.187)$	$0.953(0.212)$
	Length $= 0.046(0.005)$	$0.003(0.001)$	$0.037(0.007)$	$0.011(0.003)$
	LogTime $= -8.335(0.153)$	$-7.139(0.113)$	$-7.212(0.104)$	$-7.153(0.338)$
Norm. Approx.	$0.967(0.180)$	$0.946(0.226)$	$0.956(0.206)$	$0.945(0.229)$
	$0.046(0.005)$	$0.003(0.001)$	$0.037(0.007)$	$0.011(0.003)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.824(0.381)$	$0.848(0.359)$	$0.840(0.367)$	$0.852(0.355)$
	$0.031(0.005)$	$0.003(0.001)$	$0.027(0.006)$	$0.009(0.003)$
	$-2.588(0.008)$	$-2.107(0.084)$	$-2.123(0.009)$	$-1.027(0.008)$
Green and Shalizi [61]	$0.952(0.214)$	$0.936(0.245)$	$0.940(0.238)$	$0.910(0.286)$
	$0.044(0.007)$	$0.004(0.001)$	$0.035(0.008)$	$0.010(0.003)$
	$-1.159(0.010)$	$0.500(0.039)$	$0.199(0.045)$	$0.341(0.021)$
Levin and Levina [93]	$0.972(0.165)$	$0.966(0.181)$	$0.966(0.181)$	$0.962(0.191)$
	$0.047(0.007)$	$0.004(0.001)$	$0.040(0.009)$	$0.012(0.004)$
	$-1.148(0.005)$	$0.521(0.036)$	$0.220(0.027)$	$0.444(0.009)$

TABLE 10
Performance measures of 95% confidence intervals
$n = 80$, $\rho_n \asymp n^{-1/2}$, *graphon: smooth graphon*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.967(0.179)$	$0.931(0.253)$	$0.956(0.205)$	$0.930(0.256)$
	Length $= 0.042(0.005)$	$0.002(0.001)$	$0.026(0.005)$	$0.006(0.002)$
	LogTime $= -8.213(0.047)$	$-7.618(0.111)$	$-7.152(0.107)$	$-7.147(0.318)$
Norm. Approx.	$0.963(0.189)$	$0.932(0.252)$	$0.948(0.223)$	$0.926(0.262)$
	$0.042(0.005)$	$0.002(0.001)$	$0.026(0.005)$	$0.006(0.002)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.824(0.381)$	$0.872(0.334)$	$0.834(0.372)$	$0.852(0.355)$
	$0.029(0.004)$	Inf(NaN)	$0.021(0.005)$	$0.007(0.002)$
	$-2.575(0.006)$	$-2.185(0.026)$	$-2.114(0.018)$	$-0.854(0.036)$
Green and Shalizi [61]	$0.950(0.218)$	$0.958(0.201)$	$0.940(0.238)$	$0.920(0.272)$
	$0.041(0.006)$	$0.002(0.001)$	$0.025(0.006)$	$0.006(0.002)$
	$-1.154(0.010)$	$0.497(0.101)$	$0.176(0.039)$	$0.465(0.027)$
Levin and Levina [93]	$0.956(0.205)$	$0.974(0.159)$	$0.960(0.196)$	$0.968(0.176)$
	$0.044(0.006)$	$0.002(0.001)$	$0.029(0.007)$	$0.008(0.003)$
	$-1.150(0.006)$	$0.487(0.067)$	$0.181(0.027)$	$0.440(0.027)$

[3] AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research* **9** 1981–2014.

[4] AL HASAN, M. and DAVE, V. S. (2018). Triangle counting in large networks: a review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **8** e1226.

[5] ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis* **11** 581–598.

[6] ALI, W., WEGNER, A. E., GAUNT, R. E., DEANE, C. M. and REINERT, G. (2016). Comparison of large networks with sub-sampling strategies. *Scientific Reports* **6** 28955.

[7] ALON, U. (2007). Network motifs: theory and experimental approaches. *Nature Reviews Genetics* **8** 450–461.

[8] AMBROISE, C. and MATIAS, C. (2012). New consistent and asymptotically normal parameter estimates for random-graph mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74** 3–35.

[9] ANGST, J. and POLY, G. (2017). A weak Cramér condition and application to Edgeworth expansions. *Electronic Journal of Probability* **22**.

[10] BABU, G. J. and SINGH, K. (1984). On one term Edgeworth correction by Efron's bootstrap. *Sankhyā: The Indian Journal of Statistics, Series A* **46** 219–232.

TABLE 11
Performance measures of 95% confidence intervals
$n = 80$, $\rho_n \asymp n^{-1/2}$, *graphon: non-smooth graphon*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.974(0.159)$	$0.974(0.159)$	$0.974(0.160)$	$0.973(0.164)$
	Length $= 0.059(0.005)$	$0.011(0.002)$	$0.087(0.009)$	$0.045(0.006)$
	LogTime $= -8.196(0.051)$	$-7.297(0.134)$	$-7.314(0.140)$	$-7.011(0.405)$
Norm. Approx.	$0.973(0.162)$	$0.969(0.174)$	$0.973(0.164)$	$0.970(0.171)$
	$0.059(0.005)$	$0.011(0.002)$	$0.087(0.009)$	$0.045(0.006)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.850(0.357)$	$0.852(0.355)$	$0.852(0.355)$	$0.860(0.347)$
	$0.040(0.005)$	$0.008(0.001)$	$0.060(0.008)$	$0.033(0.006)$
	$-2.543(0.115)$	$-2.118(0.022)$	$-2.195(0.021)$	$-0.965(0.047)$
Green and Shalizi [61]	$0.944(0.230)$	$0.942(0.234)$	$0.950(0.218)$	$0.946(0.226)$
	$0.055(0.006)$	$0.011(0.002)$	$0.082(0.011)$	$0.041(0.008)$
	$-1.135(0.058)$	$0.590(0.089)$	$0.134(0.041)$	$0.383(0.038)$
Levin and Levina [93]	$0.966(0.181)$	$0.972(0.165)$	$0.968(0.176)$	$0.968(0.176)$
	$0.059(0.006)$	$0.012(0.002)$	$0.090(0.011)$	$0.048(0.008)$
	$-1.130(0.059)$	$0.556(0.029)$	$0.121(0.036)$	$0.498(0.024)$

TABLE 12
Performance measures of 95% confidence intervals
$n = 80$, $\rho_n \asymp n^{-1}$, *graphon: block model*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.987(0.115)$	$0.000(0.000)$	$0.949(0.219)$	$0.646(0.478)$
	Length $= 0.016(0.002)$	$0.000(0.000)$	$0.003(0.001)$	$0.000(0.000)$
	LogTime $= -8.487(0.176)$	$-7.230(0.105)$	$-7.422(0.114)$	$-7.183(0.322)$
Norm. Approx.	$0.983(0.129)$	$0.708(0.455)$	$0.959(0.199)$	$0.914(0.280)$
	$0.016(0.002)$	$0.000(0.000)$	$0.003(0.001)$	$0.000(0.000)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.904(0.295)$	$0.620(0.486)$	$0.898(0.303)$	$0.906(0.292)$
	$0.011(0.002)$	Inf(NaN)	Inf(NaN)	Inf(NaN)
	$-2.579(0.005)$	$-2.121(0.050)$	$-2.137(0.015)$	$-1.080(0.006)$
Green and Shalizi [61]	$0.972(0.165)$	$0.672(0.470)$	$0.896(0.306)$	$0.768(0.423)$
	$0.015(0.002)$	Inf(NaN)	$0.002(0.001)$	$0.000(0.000)$
	$-1.158(0.005)$	$0.489(0.050)$	$0.203(0.030)$	$0.299(0.015)$
Levin and Levina [93]	$0.984(0.126)$	$0.706(0.456)$	$0.996(0.063)$	$0.996(0.063)$
	$0.019(0.024)$	$0.000(0.000)$	Inf(NaN)	Inf(NaN)
	$-1.146(0.004)$	$0.509(0.031)$	$0.223(0.026)$	$0.434(0.006)$

[11] BABU, G. J. and SINGH, K. (1985). Edgeworth expansions for sampling without replacement from finite populations. *Journal of Multivariate Analysis* **17** 261–278.

[12] BABU, G. J. and SINGH, K. (1989). On Edgeworth expansions in the mixture cases. *The Annals of Statistics* **17** 443–447.

[13] BENTKUS, V., GOTZE, F. and ZITIKIS, R. (1994). Lower estimates of the convergence rate for U-statistics. *The Annals of Probability* **22** 1707–1714.

[14] BENTKUS, V., GOTZE, F. and VAN ZWET, W. R. (1997). An Edgeworth expansion for symmetric statistics. *The Annals of Statistics* 851–896.

[15] BENTKUS, V., JING, B.-Y. and ZHOU, W. (2009). On normal approximations to U-statistics. *The Annals of Probability* **37** 2174–2199.

[16] BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *The Annals of Statistics* **6** 434–451.

[17] BHATTACHARYYA, S. and BICKEL, P. J. (2015). Subsampling bootstrap of count features of networks. *The Annals of Statistics* **43** 2384–2411.

[18] BICKEL, P. J. (1974). Edgeworth Expansions in Nonparametric Statistics. *The Annals of Statistics* **2** 1–20.

[19] BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan modularities. *Proceedings of the National Academy of Sciences* **106** 21068–21073.

TABLE 13
Performance measures of 95% confidence intervals
$n = 80$, $\rho_n \asymp n^{-1}$, graphon: smooth graphon

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.984(0.126)$	$0.000(0.000)$	$0.909(0.287)$	$0.503(0.500)$
	Length $= 0.014(0.002)$	$0.000(0.000)$	$0.002(0.001)$	$0.000(0.000)$
	LogTime $= -8.205(0.054)$	$-7.243(0.071)$	$-7.276(0.103)$	$-7.098(0.400)$
Norm. Approx.	$0.981(0.138)$	$0.426(0.495)$	$0.948(0.223)$	$0.875(0.331)$
	$0.014(0.002)$	$0.000(0.000)$	$0.002(0.001)$	$0.000(0.000)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.910(0.288)$	$0.300(0.461)$	$0.910(0.288)$	$0.920(0.273)$
	$0.009(0.001)$	Inf(NaN)	Inf(NaN)	Inf(NaN)
	$-1.632(0.005)$	$-1.182(0.044)$	$-1.181(0.041)$	$-0.141(0.009)$
Green and Shalizi [61]	$0.940(0.239)$	$0.380(0.488)$	$0.900(0.302)$	$0.750(0.435)$
	$0.013(0.002)$	Inf(NaN)	$0.001(0.001)$	$0.000(0.000)$
	$-0.217(0.010)$	$1.553(0.030)$	$1.142(0.042)$	$1.167(0.016)$
Levin and Levina [93]	$0.980(0.141)$	$0.380(0.488)$	$0.970(0.171)$	$0.990(0.100)$
	$41.865(418.438)$	Inf(NaN)	Inf(NaN)	Inf(NaN)
	$-0.213(0.008)$	$1.567(0.019)$	$1.120(0.018)$	$1.245(0.014)$

TABLE 14
Performance measures of 95% confidence intervals
$n = 80$, $\rho_n \asymp n^{-1}$, graphon: non-smooth graphon

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.989(0.103)$	$0.896(0.305)$	$0.980(0.139)$	$0.911(0.285)$
	Length $= 0.022(0.002)$	$0.000(0.000)$	$0.007(0.001)$	$0.001(0.000)$
	LogTime $= -8.242(0.096)$	$-7.355(0.085)$	$-7.356(0.088)$	$-7.101(0.343)$
Norm. Approx.	$0.989(0.106)$	$0.963(0.189)$	$0.980(0.142)$	$0.962(0.192)$
	$0.022(0.002)$	$0.000(0.000)$	$0.007(0.001)$	$0.001(0.000)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.894(0.308)$	$0.908(0.289)$	$0.916(0.278)$	$0.908(0.289)$
	$0.015(0.002)$	Inf(NaN)	$0.005(0.001)$	$0.001(0.000)$
	$-2.571(0.015)$	$-2.171(0.012)$	$-2.128(0.020)$	$-1.043(0.007)$
Green and Shalizi [61]	$0.964(0.186)$	$0.968(0.176)$	$0.936(0.245)$	$0.848(0.359)$
	$0.020(0.002)$	Inf(NaN)	$0.006(0.001)$	$0.001(0.000)$
	$-1.191(0.012)$	$0.567(0.120)$	$0.219(0.024)$	$0.343(0.012)$
Levin and Levina [93]	$0.986(0.118)$	$0.984(0.126)$	$0.992(0.089)$	$0.992(0.089)$
	$0.023(0.003)$	$0.000(0.000)$	$0.008(0.002)$	$0.001(0.000)$
	$-1.183(0.010)$	$0.529(0.083)$	$0.236(0.034)$	$0.450(0.013)$

[20] BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *The Annals of Statistics* **39** 2280–2301.
[21] BICKEL, P. J., GOTZE, F. and VAN ZWET, W. (1986). The Edgeworth Expansion for U-Statistics of Degree Two. *The Annals of Statistics* **14** 1463–1484.
[22] BLOM, G. (1976). Some properties of incomplete U-statistics. *Biometrika* **63** 573–580.
[23] BLOZNELIS, M. (2003). An Edgeworth expansion for Studentized finite population statistics. *Acta Applicandae Mathematica* **78** 51–60.
[24] BLOZNELIS, M. and GÖTZE, F. (2001). Orthogonal decomposition of finite population statistics and its applications to distributional asymptotics. *The Annals of Statistics* **29** 899–917.
[25] BLOZNELIS, M. and GÖTZE, F. (2002). An Edgeworth expansion for symmetric finite population statistics. *The Annals of Probability* **30** 1238–1265.
[26] BORGS, C., CHAYES, J. and LOVÁSZ, L. (2010). Moments of two-variable functions and the uniqueness of graph limits. *Geometric And Functional Analysis* **19** 1597–1619.
[27] BROIDO, A. D. and CLAUSET, A. (2019). Scale-free networks are rare. *Nature communications* **10** 1–10.
[28] CAI, T. T. and LI, X. (2015). Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *The Annals of Statistics* **43** 1027–1059.

TABLE 15
Performance measures of 95% confidence intervals
$n = 160$, $\rho_n \asymp n^{-1/4}$, *graphon: block model*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.957(0.204)$	$0.954(0.211)$	$0.954(0.210)$	$0.951(0.216)$
	Length $= 0.048(0.004)$	$0.010(0.001)$	$0.070(0.008)$	$0.036(0.006)$
	LogTime $= -7.068(0.086)$	$-6.643(0.137)$	$-6.161(0.337)$	$-6.132(0.241)$
Norm. Approx.	$0.954(0.209)$	$0.943(0.232)$	$0.949(0.221)$	$0.943(0.232)$
	$0.048(0.004)$	$0.010(0.001)$	$0.070(0.008)$	$0.036(0.006)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.828(0.378)$	$0.834(0.372)$	$0.828(0.378)$	$0.836(0.371)$
	$0.033(0.003)$	$0.007(0.001)$	$0.049(0.007)$	$0.026(0.004)$
	$-1.198(0.004)$	$0.547(0.042)$	$0.138(0.079)$	$0.328(0.021)$
Green and Shalizi [61]	$0.934(0.249)$	$0.940(0.238)$	$0.940(0.238)$	$0.940(0.238)$
	$0.047(0.005)$	$0.010(0.002)$	$0.069(0.010)$	$0.035(0.006)$
	$0.574(0.006)$	$2.077(0.047)$	$2.548(0.041)$	$2.099(0.005)$
Levin and Levina [93]	$0.948(0.222)$	$0.952(0.214)$	$0.952(0.214)$	$0.952(0.214)$
	$0.048(0.005)$	$0.010(0.002)$	$0.070(0.010)$	$0.036(0.006)$
	$0.582(0.005)$	$2.096(0.042)$	$2.541(0.039)$	$2.268(0.005)$

TABLE 16
Performance measures of 95% confidence intervals
$n = 160$, $\rho_n \asymp n^{-1/4}$, *graphon: smooth graphon*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.958(0.200)$	$0.949(0.220)$	$0.954(0.209)$	$0.951(0.215)$
	Length $= 0.045(0.003)$	$0.005(0.001)$	$0.049(0.006)$	$0.020(0.004)$
	LogTime $= -7.305(0.064)$	$-6.596(0.222)$	$-6.151(0.328)$	$-6.117(0.240)$
Norm. Approx.	$0.954(0.209)$	$0.941(0.236)$	$0.946(0.225)$	$0.941(0.235)$
	$0.045(0.003)$	$0.005(0.001)$	$0.049(0.006)$	$0.020(0.004)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.850(0.357)$	$0.856(0.351)$	$0.842(0.365)$	$0.860(0.347)$
	$0.032(0.003)$	$0.004(0.001)$	$0.036(0.005)$	$0.016(0.003)$
	$-1.151(0.003)$	$0.512(0.045)$	$0.126(0.107)$	$0.314(0.021)$
Green and Shalizi [61]	$0.948(0.222)$	$0.946(0.226)$	$0.944(0.230)$	$0.950(0.218)$
	$0.044(0.004)$	$0.005(0.001)$	$0.049(0.007)$	$0.020(0.004)$
	$0.624(0.013)$	$2.004(0.050)$	$2.533(0.040)$	$2.103(0.010)$
Levin and Levina [93]	$0.956(0.205)$	$0.956(0.205)$	$0.964(0.186)$	$0.970(0.171)$
	$0.046(0.004)$	$0.006(0.001)$	$0.051(0.007)$	$0.022(0.004)$
	$0.625(0.009)$	$2.036(0.036)$	$2.536(0.040)$	$2.260(0.009)$

[29] CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388.
[30] CALLAERT, H. and JANSSEN, P. (1978). The Berry-Esseen Theorem for U-statistics. *The Annals of Statistics* **6** 417–421.
[31] CALLAERT, H., JANSSEN, P. and VERAVERBEKE, N. (1980). An Edgeworth Expansion for U-Statistics. *The Annals of Statistics* **8** 299–312.
[32] CALLAERT, H. and VERAVERBEKE, N. (1981). The Order of the Normal Approximation for a Studentized U-Statistic. *The Annals of Statistics* **9** 194–200.
[33] CANDES, E. J. and PLAN, Y. (2010). Matrix completion with noise. *Proceedings of the IEEE* **98** 925–936.
[34] CHAN, S. and AIROLDI, E. (2014). A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning* 208–216.
[35] CHATTERJEE, S. (2015). Matrix estimation by universal singular value thresholding. *The Annals of Statistics* **43** 177–214.
[36] CHEN, X. and KATO, K. (2019). Randomized incomplete U-statistics in high dimensions. *The Annals of Statistics* **47** 3127–3156.
[37] CHEN, S. and ONNELA, J.-P. (2019). A bootstrap method for goodness of fit and model Selection with a Single observed network. *Scientific reports* **9** 1–12.

TABLE 17
Performance measures of 95% confidence intervals
$n = 160$, $\rho_n \asymp n^{-1/4}$, *graphon: non-smooth graphon*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.959(0.198)$	$0.961(0.194)$	$0.961(0.195)$	$0.962(0.192)$
	Length $= 0.058(0.003)$	$0.034(0.002)$	$0.150(0.007)$	$0.134(0.008)$
	LogTime $= -7.164(0.080)$	$-6.145(0.477)$	$-6.043(0.343)$	$-5.933(0.330)$
Norm. Approx.	$0.958(0.201)$	$0.958(0.201)$	$0.959(0.198)$	$0.960(0.196)$
	$0.058(0.003)$	$0.034(0.002)$	$0.150(0.007)$	$0.134(0.008)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.830(0.376)$	$0.854(0.353)$	$0.840(0.367)$	$0.854(0.353)$
	$0.040(0.004)$	$0.024(0.002)$	$0.104(0.009)$	$0.093(0.009)$
	$-1.160(0.005)$	$0.489(0.054)$	$0.158(0.046)$	$0.339(0.018)$
Green and Shalizi [61]	$0.938(0.241)$	$0.936(0.245)$	$0.936(0.245)$	$0.946(0.226)$
	$0.056(0.005)$	$0.033(0.003)$	$0.145(0.013)$	$0.130(0.012)$
	$0.640(0.011)$	$2.058(0.067)$	$2.727(0.036)$	$2.164(0.022)$
Levin and Levina [93]	$0.952(0.214)$	$0.952(0.214)$	$0.954(0.210)$	$0.952(0.214)$
	$0.058(0.005)$	$0.034(0.003)$	$0.150(0.013)$	$0.135(0.012)$
	$0.640(0.013)$	$2.059(0.060)$	$2.727(0.037)$	$2.345(0.015)$

TABLE 18
Performance measures of 95% confidence intervals
$n = 160$, $\rho_n \asymp n^{-1/2}$, *graphon: block model*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.969(0.173)$	$0.964(0.186)$	$0.965(0.183)$	$0.962(0.192)$
	Length $= 0.022(0.002)$	$0.001(0.000)$	$0.012(0.002)$	$0.003(0.000)$
	LogTime $= -7.301(0.074)$	$-6.859(0.116)$	$-6.461(0.416)$	$-6.281(0.245)$
Norm. Approx.	$0.966(0.182)$	$0.960(0.195)$	$0.961(0.195)$	$0.954(0.210)$
	$0.022(0.002)$	$0.001(0.000)$	$0.012(0.002)$	$0.003(0.000)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.836(0.371)$	$0.852(0.355)$	$0.830(0.376)$	$0.864(0.343)$
	$0.015(0.002)$	$0.001(0.000)$	$0.009(0.001)$	$0.002(0.000)$
	$-1.155(0.006)$	$0.555(0.056)$	$0.192(0.114)$	$0.360(0.038)$
Green and Shalizi [61]	$0.948(0.222)$	$0.944(0.230)$	$0.938(0.241)$	$0.916(0.278)$
	$0.021(0.002)$	$0.001(0.000)$	$0.011(0.002)$	$0.002(0.000)$
	$0.624(0.008)$	$2.060(0.067)$	$2.850(0.058)$	$2.142(0.028)$
Levin and Levina [93]	$0.970(0.171)$	$0.968(0.176)$	$0.970(0.171)$	$0.968(0.176)$
	$0.022(0.002)$	$0.001(0.000)$	$0.013(0.002)$	$0.003(0.001)$
	$0.630(0.009)$	$2.073(0.057)$	$2.853(0.052)$	$2.345(0.026)$

[38] CHEN, S. X. and PENG, L. (2021+). Distributed Statistical Inference for Massive Data. *The Annals of Statistics* (To appear).

[39] CHEN, L. H. and SHAO, Q.-M. (2001). A non-uniform Berry–Esseen bound via Stein's method. *Probability theory and related fields* **120** 236–254.

[40] CHOI, D. S. (2017). Co-clustering of nonsmooth graphons. *The Annals of Statistics* **45** 1488–1515.

[41] CHOI, D. S., WOLFE, P. J. and AIROLDI, E. M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* **99** 273–284.

[42] CHUNG, F. and LU, L. (2002). Connected components in random graphs with given expected degree sequences. *Annals of combinatorics* **6** 125–145.

[43] CLAUSET, A., NEWMAN, M. E. and MOORE, C. (2004). Finding community structure in very large networks. *Physical review E* **70** 066111.

[44] CORNISH, E. A. and FISHER, R. A. (1938). Moments and cumulants in the specification of distributions. *Revue de l'Institut international de Statistique* **5** 307–320.

[45] CRAMÉR, H. (1928). On the composition of elementary errors: first paper: mathematical deductions. *Scandinavian Actuarial Journal* **1928** 13–74.

[46] CRANE, H. (2018). *Probabilistic foundations of statistical network analysis*. CRC Press.

TABLE 19
Performance measures of 95% confidence intervals
$n = 160$, $\rho_n \asymp n^{-1/2}$, *graphon: smooth graphon*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.967(0.178)$	$0.959(0.198)$	$0.964(0.186)$	$0.955(0.208)$
	Length $= 0.020(0.002)$	$0.000(0.000)$	$0.009(0.001)$	$0.001(0.000)$
	LogTime $= -7.279(0.075)$	$-6.744(0.149)$	$-6.337(0.377)$	$-6.378(0.215)$
Norm. Approx.	$0.966(0.183)$	$0.951(0.216)$	$0.958(0.201)$	$0.949(0.220)$
	$0.020(0.002)$	$0.000(0.000)$	$0.009(0.001)$	$0.001(0.000)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.852(0.355)$	$0.870(0.337)$	$0.882(0.323)$	$0.880(0.325)$
	$0.014(0.001)$	$0.000(0.000)$	$0.006(0.001)$	$0.001(0.000)$
	$-1.145(0.006)$	$0.533(0.082)$	$0.225(0.116)$	$0.452(0.012)$
Green and Shalizi [61]	$0.938(0.241)$	$0.956(0.205)$	$0.940(0.238)$	$0.920(0.272)$
	$0.019(0.002)$	$0.000(0.000)$	$0.008(0.001)$	$0.001(0.000)$
	$0.611(0.009)$	$1.997(0.116)$	$2.856(0.071)$	$2.328(0.012)$
Levin and Levina [93]	$0.962(0.191)$	$0.976(0.153)$	$0.966(0.181)$	$0.972(0.165)$
	$0.020(0.002)$	$0.000(0.000)$	$0.009(0.001)$	$0.002(0.000)$
	$0.619(0.013)$	$1.987(0.074)$	$2.848(0.055)$	$2.405(0.008)$

TABLE 20
Performance measures of 95% confidence intervals
$n = 160$, $\rho_n \asymp n^{-1/2}$, *graphon: non-smooth graphon*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.972(0.164)$	$0.974(0.159)$	$0.974(0.161)$	$0.975(0.157)$
	Length $= 0.028(0.002)$	$0.003(0.000)$	$0.029(0.002)$	$0.011(0.001)$
	LogTime $= -7.007(0.080)$	$-6.635(0.252)$	$-6.126(0.364)$	$-5.835(0.471)$
Norm. Approx.	$0.972(0.166)$	$0.973(0.163)$	$0.972(0.166)$	$0.973(0.161)$
	$0.028(0.002)$	$0.003(0.000)$	$0.029(0.002)$	$0.011(0.001)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.852(0.355)$	$0.868(0.339)$	$0.858(0.349)$	$0.874(0.332)$
	$0.018(0.002)$	$0.002(0.000)$	$0.019(0.002)$	$0.007(0.001)$
	$-1.159(0.005)$	$0.565(0.048)$	$0.149(0.076)$	$0.376(0.022)$
Green and Shalizi [61]	$0.948(0.222)$	$0.954(0.210)$	$0.956(0.205)$	$0.946(0.226)$
	$0.026(0.002)$	$0.003(0.000)$	$0.027(0.003)$	$0.009(0.001)$
	$0.636(0.019)$	$2.079(0.101)$	$2.505(0.041)$	$2.241(0.015)$
Levin and Levina [93]	$0.956(0.205)$	$0.974(0.159)$	$0.966(0.181)$	$0.974(0.159)$
	$0.028(0.002)$	$0.003(0.000)$	$0.029(0.003)$	$0.011(0.001)$
	$0.640(0.013)$	$2.062(0.050)$	$2.519(0.038)$	$2.462(0.015)$

[47] DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap methods and their application*. Cambridge university press.

[48] DELLA ROSSA, F., DERCOLE, F. and PICCARDI, C. (2013). Profiling core-periphery network structure by random walkers. *Scientific reports* **3** 1–8.

[49] DIACONIS, P. and JANSON, S. (2007). Graph limits and exchangeable random graphs. *arXiv preprint arXiv:0712.2749*.

[50] DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator. *The Annals of Mathematical Statistics* **27** 642–669.

[51] EDGEWORTH, F. (1905). The law of error. In *Cambridge Philos. Soc.* **20** 36–66 and 113–141.

[52] FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications* **1**. Wiley.

[53] FISHER, S. R. A. and CORNISH, E. (1960). The percentile points of distributions having known cumulants. *Technometrics* **2** 209–225.

[54] GAO, C. and LAFFERTY, J. (2017a). Testing network structure using relations between small subgraph probabilities. *arXiv preprint arXiv:1704.06742*.

[55] GAO, C. and LAFFERTY, J. (2017b). Testing for global network structure using small subgraph statistics. *arXiv preprint arXiv:1710.00862*.

TABLE 21
Performance measures of 95% confidence intervals
$n = 160$, $\rho_n \asymp n^{-1}$, *graphon: block model*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.989(0.106)$	$0.000(0.000)$	$0.975(0.157)$	$0.813(0.390)$
	Length $= 0.006(0.000)$	$0.000(0.000)$	$0.000(0.000)$	$0.000(0.000)$
	LogTime $= -7.284(0.063)$	$-6.939(0.150)$	$-6.462(0.321)$	$-6.293(0.247)$
Norm. Approx.	$0.988(0.109)$	$0.738(0.440)$	$0.976(0.154)$	$0.948(0.221)$
	$0.006(0.000)$	$0.000(0.000)$	$0.000(0.000)$	$0.000(0.000)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.872(0.334)$	$0.652(0.477)$	$0.912(0.284)$	$0.920(0.272)$
	$0.004(0.000)$	Inf(NaN)	$0.000(0.000)$	Inf(NaN)
	$-1.157(0.017)$	$0.495(0.039)$	$0.187(0.099)$	$0.271(0.018)$
Green and Shalizi [61]	$0.966(0.181)$	$0.708(0.455)$	$0.768(0.423)$	$0.568(0.496)$
	$0.005(0.001)$	Inf(NaN)	$0.000(0.000)$	$0.000(0.000)$
	$0.619(0.014)$	$2.030(0.038)$	$2.855(0.072)$	$2.150(0.015)$
Levin and Levina [93]	$0.980(0.140)$	$0.728(0.445)$	$0.990(0.100)$	$0.992(0.089)$
	$0.006(0.001)$	$0.000(0.000)$	$0.001(0.000)$	$0.000(0.000)$
	$0.626(0.009)$	$2.041(0.030)$	$2.874(0.060)$	$2.371(0.017)$

TABLE 22
Performance measures of 95% confidence intervals
$n = 160$, $\rho_n \asymp n^{-1}$, *graphon: smooth graphon*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.989(0.105)$	$0.000(0.000)$	$0.964(0.186)$	$0.651(0.477)$
	Length $= 0.005(0.000)$	$0.000(0.000)$	$0.000(0.000)$	$0.000(0.000)$
	LogTime $= -7.290(0.076)$	$-6.949(0.103)$	$-6.458(0.388)$	$-6.277(0.333)$
Norm. Approx.	$0.987(0.115)$	$0.437(0.496)$	$0.968(0.177)$	$0.926(0.262)$
	$0.005(0.000)$	$0.000(0.000)$	$0.000(0.000)$	$0.000(0.000)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.870(0.338)$	$0.290(0.456)$	$0.910(0.288)$	$0.900(0.302)$
	$0.003(0.000)$	Inf(NaN)	$0.000(0.000)$	Inf(NaN)
	$-0.193(0.061)$	$1.592(0.024)$	$1.132(0.038)$	$1.146(0.012)$
Green and Shalizi [61]	$0.970(0.171)$	$0.380(0.488)$	$0.760(0.429)$	$0.560(0.499)$
	$0.005(0.001)$	Inf(NaN)	$0.000(0.000)$	$0.000(0.000)$
	$1.543(0.035)$	$2.949(0.030)$	$3.809(0.034)$	$3.020(0.005)$
Levin and Levina [93]	$0.980(0.141)$	$0.390(0.490)$	$0.990(0.100)$	$0.990(0.100)$
	$0.005(0.001)$	$0.000(0.000)$	$0.000(0.000)$	$0.000(0.000)$
	$1.541(0.033)$	$2.968(0.033)$	$3.803(0.038)$	$3.069(0.009)$

[56] GAO, C., LU, Y. and ZHOU, H. H. (2015). Rate-optimal graphon estimation. *The Annals of Statistics* **43** 2624–2652.

[57] GAO, F., MA, Z. and YUAN, H. (2020). Community detection in sparse latent space models. *arXiv preprint arXiv:2008.01375*.

[58] GAO, C., LU, Y., MA, Z. and ZHOU, H. H. (2016). Optimal estimation and completion of matrices with biclustering structures. *The Journal of Machine Learning Research* **17** 5602–5630.

[59] GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2018). Community detection in degree-corrected block models. *The Annals of Statistics* **46** 2153–2185.

[60] GEL, Y. R., LYUBCHICH, V. and RAMIREZ, L. L. R. (2017). Bootstrap quantification of estimation uncertainties in network degree distributions. *Scientific reports* **7** 1–12.

[61] GREEN, A. and SHALIZI, C. R. (2017). Bootstrapping exchangeable random graphs. *arXiv preprint arXiv:1711.00813*.

[62] GROVER, A. and LESKOVEC, J. (2016). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining* 855–864.

[63] GUÉDON, O. and VERSHYNIN, R. (2016). Community detection in sparse networks via Grothendieck's inequality. *Probability Theory and Related Fields* **165** 1025–1049.

TABLE 23
Performance measures of 95% confidence intervals
$n = 160$, $\rho_n \asymp n^{-1}$, *graphon: non-smooth graphon*

Method	Edge	Triangle	V-shape	Three star
Our method	Coverage $= 0.992(0.090)$	$0.947(0.223)$	$0.987(0.112)$	$0.960(0.197)$
	Length $= 0.008(0.001)$	$0.000(0.000)$	$0.001(0.000)$	$0.000(0.000)$
	LogTime $= -7.275(0.072)$	$-6.576(0.324)$	$-6.121(0.356)$	$-6.304(0.268)$
Norm. Approx.	$0.991(0.097)$	$0.974(0.159)$	$0.984(0.127)$	$0.974(0.158)$
	$0.008(0.001)$	$0.000(0.000)$	$0.001(0.000)$	$0.000(0.000)$
	No time cost	No time cost	No time cost	No time cost
Bhattacharyya and Bickel [17]	$0.880(0.325)$	$0.926(0.262)$	$0.882(0.323)$	$0.894(0.308)$
	$0.005(0.001)$	Inf(NaN)	$0.001(0.000)$	$0.000(0.000)$
	$-1.195(0.007)$	$0.560(0.055)$	$0.249(0.058)$	$0.317(0.019)$
Green and Shalizi [61]	$0.968(0.176)$	$0.984(0.126)$	$0.892(0.311)$	$0.650(0.477)$
	$0.007(0.001)$	Inf(NaN)	$0.001(0.000)$	$0.000(0.000)$
	$0.575(0.007)$	$2.011(0.069)$	$2.893(0.040)$	$2.187(0.014)$
Levin and Levina [93]	$0.988(0.109)$	$0.994(0.077)$	$0.988(0.109)$	$0.986(0.118)$
	$0.008(0.001)$	$0.000(0.000)$	$0.001(0.000)$	$0.000(0.000)$
	$0.569(0.007)$	$1.986(0.070)$	$2.895(0.035)$	$2.401(0.014)$

[64] HAEUSLER, E. (1988). On the rate of convergence in the central limit theorem for martingales with discrete and continuous time. *The Annals of Probability* 275–299.
[65] HALL, P. (1983). Inverting an Edgeworth expansion. *The Annals of Statistics* **11** 569–576.
[66] HALL, P. (1987). Edgeworth expansion for Student's t statistic under minimal moment conditions. *The Annals of Probability* **15** 920–931.
[67] HALL, P. (1993). On Edgeworth expansion and bootstrap confidence bands in nonparametric curve estimation. *Journal of the Royal Statistical Society: Series B (Methodological)* **55** 291–304.
[68] HALL, P. (2013). *The bootstrap and Edgeworth expansion*. Springer Science & Business Media.
[69] HALL, P. and HEYDE, C. C. (2014). *Martingale limit theory and its application*. Academic press.
[70] HALL, P. and JING, B.-Y. (1995). Uniform Coverage Bounds for Confidence Intervals and Berry-Esseen Theorems for Edgeworth Expansion. *The Annals of Statistics* **23** 363–375.
[71] HELMERS, R. (1991). On the Edgeworth Expansion and the Bootstrap Approximation for a Studentized U-Statistic. *The Annals of Statistics* **19** 470–484.
[72] HEYDE, C. and BROWN, B. (1970). On the Departure from Normality of a Certain Class of Martingales. *The Annals of Mathematical Statistics* **41** 2161–2165.
[73] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* **19** 293–325.
[74] HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association* **97** 1090–1098.
[75] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Social networks* **5** 109–137.
[76] HOLLAND, P. W. and LEINHARDT, S. (1971). Transitivity in structural models of small groups. *Comparative group studies* **2** 107–124.
[77] HOOVER, D. N. (1979). Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton, NJ* **2**.
[78] HUNTER, D. R., HANDCOCK, M. S., BUTTS, C. T., GOODREAU, S. M. and MORRIS, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of statistical software* **24** nihpa54860.
[79] JANSSEN, P. (1981). Rate of convergence in the central limit theorem and in the strong law of large numbers for von Mises statistics. *Metrika* **28** 35–46.
[80] JING, B.-Y. and WANG, Q. (2003). Edgeworth expansion for U-statistics under minimal conditions. *The Annals of Statistics* **31** 1376–1391.
[81] JING, B.-Y. and WANG, Q. (2010). A unified approach to edgeworth expansions for a general class of statistics. *Statistica Sinica* **20** 613–636.
[82] JING, B.-Y., LI, T., LYU, Z. and XIA, D. (2020). Community Detection on Mixture Multi-layer Networks via Regularized Tensor Decomposition. *arXiv preprint arXiv:2002.04457*.
[83] KARRER, B. and NEWMAN, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical review E* **83** 016107.

[84] KHETAN, A. and OH, S. (2019). Spectrum estimation from a few entries. *The Journal of Machine Learning Research* **20** 718–772.
[85] KLOPP, O., TSYBAKOV, A. B. and VERZELEN, N. (2017). Oracle inequalities for network models and sparse graphon estimation. *The Annals of Statistics* **45** 316–354.
[86] KOLASSA, J. E. and MCCULLAGH, P. (1990). Edgeworth series for lattice distributions. *The Annals of Statistics* **18** 981–985.
[87] KONG, X. and ZHENG, W. ((To appear)). Design based incomplete U-statistics. *Statistica Sinica*.
[88] KOSOROK, M. R. (2007). *Introduction to empirical processes and semiparametric inference*. Springer Science & Business Media.
[89] LAHIRI, S. N. (1993). Bootstrapping the studentized sample mean of lattice variables. *Journal of multivariate analysis* **45** 247–256.
[90] LAI, T. L. and WANG, J. Q. (1993). Edgeworth expansions for symmetric statistics with applications to bootstrap methods. *Statistica Sinica* **3** 517–542.
[91] LEI, J. (2016). A goodness-of-fit test for stochastic block models. *The Annals of Statistics* **44** 401–424.
[92] LESKOVEC, J. and KREVL, A. (2014). SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data.
[93] LEVIN, K. and LEVINA, E. (2019). Bootstrapping Networks with Latent Space Structure. *arXiv preprint arXiv:1907.10821*.
[94] LI, Y. and LI, H. (2018). Two-sample test of community memberships of weighted stochastic block models. *arXiv preprint arXiv:1811.12593*.
[95] LOVÁSZ, L. (2009). Very large graphs. *Current Developments in Mathematics* **2008** 67–128.
[96] MAESONO, Y. (1997). Edgeworth expansions of a studentized U-statistic and a jackknife estimator of variance. *Journal of Statistical Planning and Inference* **61** 61–84.
[97] MAJOR, P. (2007). On a Multivariate Version of Bernstein's Inequality. *Electron. J. Probab.* **12** 966–988.
[98] MATIAS, C., SCHBATH, S., BIRMELÉ, E., DAUDIN, J.-J. and ROBIN, S. (2006). Network motifs: mean and variance for the count. *REVSTAT-Statistical Journal* **4** 31–51.
[99] MATSUSHITA, Y. and OTSU, T. (To appear). Jackknife empirical likelihood: small bandwidth, sparse network and high-dimension asymptotic. *Biometrika*.
[100] MATTNER, L. and SHEVTSOVA, I. (2017). An optimal Berry-Esseen type theorem for integrals of smooth functions. *arXiv preprint arXiv:1710.08503*.
[101] MAUGIS, P., PRIEBE, C. E., OLHEDE, S. C. and WOLFE, P. J. (2017). Statistical inference for network samples using subgraph counts. *arXiv preprint arXiv:1701.00505*.
[102] 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.
[103] NEWMAN, M. E. J. (2009). Random Graphs with Clustering. *Phys. Rev. Lett.* **103** 058701.
[104] NEWMAN, M. E. and LEICHT, E. A. (2007). Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences* **104** 9564–9569.
[105] NOWICKI, K. and WIERMAN, J. C. (1988). Subgraph counts in random graphs using incomplete U-statistics methods. *Discrete Mathematics* **72** 299–310.
[106] OUADAH, S., ROBIN, S. and LATOUCHE, P. (2019). Degree-based goodness-of-fit tests for heterogeneous random graph models: Independent and exchangeable cases. *Scandinavian Journal of Statistics* **47** 156-181.
[107] OUCHTI, L. (2005). On the rate of convergence in the central limit theorem for martingale difference sequences. In *Annales de l'IHP Probabilités et statistiques* **41** 35–43.
[108] PETROV, V. V. (2012). *Sums of independent random variables* **82**. Springer Science & Business Media.
[109] PRŽULJ, N. (2007). Biological network comparison using graphlet degree distribution. *Bioinformatics* **23** e177–e183.
[110] PUTTER, H. and VAN ZWET, W. R. (1998). Empirical Edgeworth expansions for symmetric statistics. *The Annals of Statistics* **26** 1540–1569.
[111] RIZZO, M. L. and SZÉKELY, G. J. (2016). Energy distance. *wiley interdisciplinary reviews: Computational statistics* **8** 27–38.
[112] ROHE, K. (2019). A critical threshold for design effects in network sampling. *The Annals of Statistics* **47** 556–582.
[113] ROHE, K. and QIN, T. (2013). The blessing of transitivity in sparse and stochastic networks. *arXiv preprint arXiv:1307.2302*.
[114] RUBINOV, M. and SPORNS, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52** 1059–1069.
[115] SAKOV, A. and BICKEL, P. J. (2000). An Edgeworth expansion for the m out of n bootstrapped median. *Statistics & probability letters* **49** 217–223.

[116] SCHUDY, W. and SVIRIDENKO, M. (2011). Bernstein-like concentration and moment inequalities for polynomials of independent random variables: multilinear case. *arXiv preprint arXiv:1109.5193*.

[117] SESHADHRI, C., SHARMA, A., STOLMAN, A. and GOEL, A. (2020). The impossibility of low-rank representations for triangle-rich complex networks. *Proceedings of the National Academy of Sciences*.

[118] SINGH, K. (1981). On the Asymptotic Accuracy of Efron's Bootstrap. *The Annals of Statistics* **9** 1187–1195.

[119] SONG, K. (2016). Ordering-free inference from locally dependent data. *arXiv preprint arXiv:1604.00447*.

[120] SONG, K. (2018). A Uniform-in-P Edgeworth Expansion under Weak Cramer Conditions. *arXiv preprint arXiv:1806.01431*.

[121] SONG, Y., CHEN, X. and KATO, K. (2019). Approximating high-dimensional infinite-order U-statistics: Statistical and computational guarantees. *Electronic Journal of Statistics* **13** 4794–4848.

[122] STEPHEN, A. T. and TOUBIA, O. (2009). Explaining the power-law degree distribution in a social commerce network. *Social Networks* **31** 262–270.

[123] TANG, M., ATHREYA, A., SUSSMAN, D. L., LYZINSKI, V., PARK, Y. and PRIEBE, C. E. (2017). A semiparametric two-sample hypothesis testing problem for random graphs. *Journal of Computational and Graphical Statistics* **26** 344–354.

[124] THOMPSON, M. E., RAMIREZ RAMIREZ, L. L., LYUBCHICH, V. and GEL, Y. R. (2016). Using the bootstrap for statistical inference on random graphs. *Canadian Journal of Statistics* **44** 3–24.

[125] TSIOTAS, D. (2019). Detecting different topologies immanent in scale-free networks with the same degree distribution. *Proceedings of the National Academy of Sciences* **116** 6701–6706.

[126] VERSHYNIN, R. (2017). Four lectures on probabilistic methods for data science. *The Mathematics of Data, IAS/Park City Mathematics Series* 231–271.

[127] WALLACE, D. L. (1958). Asymptotic approximations to distributions. *The Annals of Mathematical Statistics* **29** 635–654.

[128] WANG, Y. X. R. and BICKEL, P. J. (2017). Likelihood-based model selection for stochastic block models. *The Annals of Statistics* **45** 500–528.

[129] WASSERMAN, L. (2006). *All of nonparametric statistics*. Springer Science & Business Media.

[130] WATTS, D. J. and STROGATZ, S. H. (1998). Collective dynamics of 'small-world'networks. *Nature* **393** 440.

[131] WEGNER, A. E., OSPINA-FORERO, L., GAUNT, R. E., DEANE, C. M. and REINERT, G. (2018). Identifying networks with common organizational principles. *Journal of Complex Networks* **6** 887–913.

[132] WOODROOFE, M. and JHUN, M. (1988). Singh's theorem in the lattice case. *Statistics & probability letters* **7** 201–205.

[133] XIA, D. and YUAN, M. (2021). Statistical inferences of linear forms for noisy matrix completion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **83** 58-77.

[134] XU, J. (2018). Rates of Convergence of Spectral Methods for Graphon Estimation. In *Proceedings of the 35th International Conference on Machine Learning*.

[135] YAN, T., LENG, C. and ZHU, J. (2016). Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *The Annals of Statistics* **44** 31–57.

[136] ZHANG, Y., LEVINA, E. and ZHU, J. (2017). Estimating network edge probabilities by neighbourhood smoothing. *Biometrika* **104** 771–783.

[137] ZHANG, Y., LEVINA, E. and ZHU, J. (2020+). Detecting overlapping communities in networks using spectral methods. *SIAM Journal on Mathematics of Data Science* (In press).

[138] ZHANG, X., MARTIN, T. and NEWMAN, M. E. (2015). Identification of core-periphery structure in networks. *Physical Review E* **91** 032803.

[139] ZHANG, A. Y. and ZHOU, H. H. (2016). Minimax rates of community detection in stochastic block models. *The Annals of Statistics* **44** 2252–2280.

[140] ZHANG, A. Y. and ZHOU, H. H. (2020). Theoretical and computational guarantees of mean field variational inference for community detection. *The Annals of Statistics* **48** 2575–2598.

[141] ZHAO, Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics* **40** 2266–2292.