ASYMPTOTIC NORMALITY OF SUBGRAPH COUNTS IN SPARSE INHOMOGENEOUS RANDOM GRAPHS

SAYAK CHATTERJEE, ANIRBAN CHATTERJEE, ABHINAV CHAKRABORTY, AND BHASWAR B. BHATTACHARYA

Abstract. In this paper, we derive the asymptotic distribution of the number of copies of a fixed graph H in a random graph G_n sampled from a sparse graphon model. Specifically, we provide a refined analysis that separates the contributions of edge randomness and vertex-label randomness, allowing us to identify distinct sparsity regimes in which each component dominates or both contribute jointly to the fluctuations. As a result, we establish asymptotic normality for the count of any fixed graph H in G_n across the entire range of sparsity (above the containment threshold for H in G_n). These results provide a complete description of subgraph count fluctuations in sparse inhomogeneous networks, closing several gaps in the existing literature that were limited to specific motifs or suboptimal sparsity assumptions.

1. Introduction

A *graphon* is a symmetric measurable function $W : [0,1]^2 \to [0,1]$, that is, $W(x,y) = W(y,x)$ for all $x, y \in [0,1]$. Graphons arise as limit objects of sequences of large graphs and have attracted significant attention in recent years. They form a powerful bridge between combinatorics and analysis and have found applications across a range of disciplines, including statistical physics, probability, and statistics; see, for example, [4, 13, 15, 19, 20]. For a comprehensive exposition of the theory of graph limits, we refer to Lovász [41]. Graphons also provide a natural probabilistic model for generating inhomogeneous random graphs, which generalize the classical (homogeneous) Erdős–Rényi random graph, where edge probabilities vary across vertex pairs according to the underlying graphon (see [9, 12, 24, 42] among others).

Random graphs generated from a graphon model produces dense graphs, which have $O(n^2)$ edges with high probability. In contrast, many real-world networks, arising from social, biological, and technological applications, are observed to be sparse. To model such sparsity, the *sparse graphon model* has been proposed, which introduces a sparsity parameter ρ_n for controlling the edge density (see [7, 8, 36, 43] references therein):

Definition 1.1 (Sparse graphon model)**.** Given a sparsity parameter $\rho_n \in (0,1)$ and a graphon W with $\int_{[0,1]^2} W(x,y)\mathrm{d}x\mathrm{d}y > 0$, the *$W$-random graph with sparsity ρ_n* on the vertex set $[n] := \{1, 2, \ldots, n\}$, hereafter denoted by $G(n, \rho_n, W)$, is obtained by connecting the vertices i and j with probability $\rho_n W(U_i, U_j)$, independently for all $1 \leqslant i < j \leqslant n$, where $\{U_i : 1 \leqslant i \leqslant n\}$ is a sequence of i.i.d. $U[0,1]$ random variables.

The model in Definition 1.1 encompasses many well-known network models, including the classical Erdős-Rényi random graph model (where W is a constant function), the stochastic block model [7, 31] (where W is a block function), latent space models [29], random dot-product graphs [1, 49], and random geometric graphs [47], among others. Over the years, the sparse graphon model has emerged as a fundamental tool in modern network analysis, with applications in community detection, subgraph count statistics, and nonparametric estimation of graph parameters, among others.

Among the basic features of a network are its subgraph (motif) counts, that is, the frequencies of particular patterns (subgraphs), such as the number or density of edges, triangles, or stars within the network. These encode important structural information about the geometry of a network, and many features of practical interest can be derived from motif counts, such as the clustering coefficient [55], degree distribution [48], and transitivity [30] (see [50] for others). In the framework of the graphon model, subgraph counts are often referred to as network moments [8, 14], which provide an important tool for inferring properties of the underlying graphon (see [6, 27, 38, 43, 45, 54] and references therein). Consequently, understanding the asymptotic properties of subgraph counts in W-random graphs is a problem of central importance in statistical network analysis. To this end, given a fixed graph $H = (V(H), E(H))$ denote by $X_n(H, G_n)$ the number of copies of H in the W-random graph $G(n, \rho_n, W)$. More formally,

$$X_n(H, G_n) = \sum_{1 \leqslant i_1 < \cdots < i_{|V(H)|} \leqslant n} \; \sum_{H' \in \mathscr{G}_H(\{i_1, \ldots, i_{|V(H)|}\})} \; \prod_{(i_s, i_t) \in E(H')} a_{i_s i_t}, \tag{1.1}$$

where, for any set $S \subseteq [n]$, $\mathscr{G}_H(S)$ denotes the collection of all subgraphs of the complete graph $K_{|S|}$ on the vertex set S which are isomorphic to H.[1] Note that

$$|\mathscr{G}_H(\{1, \ldots, |V(H)|\})| = \frac{|V(H)|!}{|\mathrm{Aut}(H)|}, \tag{1.2}$$

where $\mathrm{Aut}(H)$ is the set of all automorphisms of the graph H, that is, the collection of permutations of the vertex set $V(H)$ such that $(x, y) \in E(H)$ if and only if $(\sigma(x), \sigma(y)) \in E(H)$. Hence,

$$\begin{aligned} \mathbb{E}[X(H, G_n)] &= \sum_{1 \leqslant i_1 < \cdots < i_{|V(H)|} \leqslant n} \; \sum_{H' \in \mathscr{G}_H(\{i_1, \ldots, i_{|V(H)|}\})} \; \prod_{(i_s, i_t) \in E(H')} \rho_n \mathbb{E}[W(U_{i_s} U_{i_t})] \\ &= \frac{(n)_{|V(H)|}}{|\mathrm{Aut}(H)|} \rho_n^{|E(H)|} t(H, W), \end{aligned} \tag{1.3}$$

where $(n)_{|V(H)|} := n(n-1) \cdots (n - |V(H)| + 1)$ and

$$t(H, W) = \int_{[0,1]^{|V(H)|}} \prod_{(a,b) \in E(H)} W(x_a, x_b) \prod_{a=1}^{|V(H)|} \mathrm{d}x_a \tag{1.4}$$

is the *homomorphism density* of the graph H in the graphon W.

Classical results of Barbour et al. [3], Janson [33], Nowicki [46], Ruciński [51] provide a complete characterization of the asymptotic normality of $X(H, G_n)$ (after appropriate centering and scaling) when $G_n \sim G(n, \rho_n)$ is the (homogeneous) Erdős-Rényi model (see [34, Chapter 6] for comprehensive treatment). For $G_n \sim G(n, \rho_n, W)$ sampled from the sparse random graphon model, the fluctuation of $X(H, G_n)$ was first studied in the seminal paper of Bickel et al. [8]. This result has been refined and extended in many directions (see the discussion Section 2.4). However, existing results remain incomplete: they apply only to specific types of subgraphs and often do not cover the full sparsity regime. Our objective this paper is to fill these missing gaps. Specifically, we establish the asymptotic normality of $Z(H, G_n)$ for all fixed graphs H in the entire allowable sparsity regime (that is, whenever ρ_n is above the containment threshold of H in $G(n, \rho_n, W)$). In fact, we provide a more fine-grained analysis that decouples the randomness of the vertex labels from the randomness of the edges and identifies distinct regimes of sparsity

[1] Note that we count unlabelled copies of H. Several other papers count labelled copies, where $X(H, G_n)$ is multiplied by $|\mathrm{Aut}(H)|$.

where the two components contribute, either separately or simultaneously, to the fluctuations of $Z(H, G_n)$. This relies on decomposing the centered subgraph count as follows:

$$\begin{aligned}\Delta(H, G_n) &:= X(H, G_n) - \mathbb{E}[X(H, G_n)] \\ &= \big(X(H, G_n) - \mathbb{E}[X(H, G_n) \mid \mathcal{F}(\boldsymbol{U}_n)]\big) + \big(\mathbb{E}[X(H, G_n) \mid \mathcal{F}(\boldsymbol{U}_n)] - \mathbb{E}[X(H, G_n)]\big) \\ &=: \Delta_1(H, G_n) + \Delta_2(H, G_n), \end{aligned} \tag{1.5}$$

where $\mathcal{F}(\boldsymbol{U}_n)$ is the sigma-algebra generated by the collection $\boldsymbol{U}_n = \{U_1, U_2, \ldots, U_n\}$. Note that, conditioned on $\boldsymbol{U}_n$, the randomness in the first term $\Delta_1(H, G_n)$ arises solely from the edges of G_n. On the other hand, $\Delta_2(H, G_n)$ depends only on vertex labels $\{U_1, U_2, \ldots, U_n\}$, that is, it is $\mathcal{F}(\boldsymbol{U}_n)$-measurable. Contributions of $\Delta_1(H, G_n)$ and $\Delta_2(H, G_n)$ to the total variance of $X(H, G_n)$ depend on the regime of sparsity and also on a notion of H-regularity of the graphon W (see Definition 2.4). Specifically, we obtain the following results:

- If W is H-regular, the contribution of $\Delta_2(H, G_n)$ to the variance of $X(H, G_n)$ is asymptotically negligible, and the limiting distribution of $X(H, G_n)$ is governed solely by $\Delta_1(H, G_n)$ (see Theorem 2.6).
- If W is not H-regular, the asymptotic behaviors of $\Delta_1(H, G_n)$ and $\Delta_2(H, G_n)$ exhibits a phase transition. Below a critical threshold, $\Delta_2(H, G_n)$ contributes negligibly to the variance of $X(H, G_n)$, and the fluctuations are driven entirely by $\Delta_1(H, G_n)$. Above the threshold, the roles reverse, with $\Delta_1(H, G_n)$ becoming negligible and $\Delta_2(H, G_n)$ governing the fluctuations. At the threshold, both terms contribute, and the appropriately standardized pair $(\Delta_1(H, G_n), \Delta_2(H, G_n))$ converges jointly to independent normal distributions with nontrivial variances (see Theorem 2.11).

The above result combined show that for any subgraph $H = (V(H), E(H))$, with $|E(H)| \geqslant 1$, $Z(H, G_n)$ is asymptotically normal whenever ρ_n is above the containment threshold of H in $G(n, W)$ (see Corollary 2.13). This completes the program, initiated Bickel et al. [8], of deriving the asymptotic distribution of subgraph counts (network moments) in sparse graphon models in full generality. In particular, it also answers a question posed by Hladkỳ et al. [28] about the distribution of the number of cliques. As a byproduct of our analysis we obtain a conditional central limit theorem for $\Delta_1(H, G_n)$ (in the Wasserstein distance) that might be of independent interest (see Proposition 2.15).

Asymptotic Notations. Throughout the paper we will use the following asymptotic notations: For two sequences a_n and b_n we will write $a_n \lesssim b_n$ and $a_n = O(b_n)$ if for all n large enough $a_n \leqslant C_1 b_n$, for some constant $C_1 > 0$. Similarly, $a_n \gtrsim b_n$ will mean $a_n \geqslant C_2 b_n$ and $a_n \asymp b_n$ will mean $C_2 b_n \leqslant a_n \leqslant C_1 b_n$, for n large enough and constants $C_1, C_2 > 0$. Subscripts in the above notation, for example $\lesssim_{\square}$, $O_{\square}$, $\gtrsim_{\square}$, and $\asymp_{\square}$ denote that the hidden constants may depend on the subscripted parameters. Moreover, $a_n \ll b_n$, $a_n \gg b_n$, and $a_n \sim b_n$ will mean $a_n/b_n \to 0$, $a_n/b_n \to \infty$, $a_n/b_n \to 1$, as $n \to \infty$.

2. Fluctuations of Subgraph Counts in Sparse W-Random Graphs

In this section we will state our main result on the asymptotic distribution $Z(H, G_n)$, where $G_n \sim G(n, \rho_n, W)$ is a W-random graph with sparsity ρ_n. The section is organized as follows.

2.1. **Preliminaries.** The preliminary towards deriving the asymptotic distribution is $Z(H, G_n)$ is to figure out rate at which ρ_n has to diverge such that $G_n \sim G(n, \rho_n, W)$ contains at least one (or infinitely many) copies of H. For the classical Erdős-Rényi model this is standard fare in the random graphs literature (see [34, Chapter 3]). In particular, it depends on whether or not the graph H is balanced:

FIGURE 1. (a) A graph which is balanced, but not strictly balanced; (b) an unbalanced graph.

Definition 2.1 (Balanced graph)**.** [34, Equation (3.6)] For a fixed graph $H = (V(H), E(H))$, define

$$m(H) := \max_{F \subseteq H, |V(F)| \geqslant 1} \frac{|E(F)|}{|V(F)|}, \tag{2.1}$$

where the maximum is taken over all possible non-empty subgraphs of H. If the maximum is attained by H itself, that is, $m(H) = \frac{|E(H)|}{|V(H)|}$, then H said to be a *balanced* graph. Otherwise, H said to be an *unbalanced* graph. Moreover, if the maximum in (2.1) uniquely attained H, then H is called a *strictly balanced* graph.

Examples of strictly balanced graphs include trees, cycles, and complete graphs. Figure 1 shows examples for the other cases:

- Figure 1 (a) is an example of a graph H which is balanced, but not strictly balanced. Here, (2.1) is attained both at the full graph H as well as the subgraph F marked by the red dotted circle.
- Figure 1 (b) is an example of an unbalanced graph H. Here, for the subgraph F marked by the red dotted circle, $\frac{|E(F)|}{|V(F)|} = \frac{5}{4} > \frac{6}{5} = \frac{|E(H)|}{|V(H)|}$.

It is well-known from classical results [10, 25] that for an Erdős-Rényi random $G(n, \rho_n)$, the threshold for the appearance of a fixed graph $H = (V(H), E(H))$ is determined by the condition $n\rho_n^{m(H)} \asymp 1$ (see [34, Theorem 3.4]). The proof of this result can be easily adapted to show that the threshold remains the same for the inhomogeneous random graph model $G(n, \rho, W)$ as well. We summarize this result in the following proposition. A proof is included in Section 3 for the sake of completeness.

Proposition 2.2. *Fix a graph $H = (V(H), E(H))$ with $|E(H)| \geqslant 1$ and a graphon W such that $t(H, W) > 0$. Then for $G_n \sim G(n, \rho_n, W)$ the following holds:*

$$X(H, G_n) \xrightarrow{P} \begin{cases} 0 & \text{if } n\rho_n^{m(H)} \ll 1, \\ \infty & \text{if } n\rho_n^{m(H)} \gg 1. \end{cases}$$

2.2. **Statements of the Results.** In light of the above proposition, hereafter, we will assume the following:

Assumption 2.3. Throughout, we will assume the following on ρ_n, H, and W:

- $\rho_n \in (0, 1)$ is such that $n^{-\frac{1}{m(H)}} \ll \rho_n \ll 1$.
- $H = (V(H), E(H))$ is a fixed graph with $|E(H)| \geqslant 1$.
- W is a graphon with $t(H, W) > 0$.

Now, define

$$Z(H, G_n) = \frac{X(H, G_n) - \mathbb{E}[X(H, G_n)]}{\sqrt{\mathrm{Var}[X(H, G_n)]}} = \frac{\Delta(H, G_n)}{\mathrm{Var}[\Delta(H, G_n)]}, \tag{2.2}$$

where $\Delta(H, G_n) := X(H, G_n) - \mathbb{E}[X(H, G_n)]$. Recalling the the decomposition in (2.11), note that

$$\begin{aligned}\mathrm{Var}[\Delta(H, G_n)] &= \mathbb{E}[\mathrm{Var}[\Delta(H, G_n) \mid \mathcal{F}(\boldsymbol{U}_n)]] + \mathrm{Var}[\mathbb{E}[\Delta(H, G_n) \mid \mathcal{F}(\boldsymbol{U}_n)]] \\ &= \mathrm{Var}[\Delta_1(H,\, G_n)] + \mathrm{Var}[\Delta_2(H,\, G_n)]. \end{aligned} \tag{2.3}$$

Our objective is to identify the regimes in $\Delta_1(H, G_n)$ and $\Delta_2(H, G_n)$ contribute to the total variance. This depends on the regime of sparsity and also on how H is embedded in W. To this end, we need the following definition:

Definition 2.4 (H-regularity)**.** A graphon W is said to be H-regular if, for almost every $x \in [0,\, 1]$,

$$\bar{t}(x,\, H,\, W) := \frac{1}{|V(H)|} \sum_{a=1}^{|V(H)|} t_a(x, H, W) = t(H, W), \tag{2.4}$$

where

$$t_a(x, H, W) := \mathbb{E}\left[\prod_{(s,\, t) \in E(H)} W(U_s,\, U_t) \mid U_a = x\right],$$

for $1 \leqslant a \leqslant |V(H)|$.

Note that in (2.4) it suffices assume that $\bar{t}(x, H, W)$ is a constant for almost every $x \in [0, 1]$. This is because

$$\int_0^1 t_a(x, H, W)\mathrm{d}x = t(H, W),$$

for all $a \in V(H)$. Hence, if $\bar{t}(x, H, W)$ is a constant almost everywhere, then the constant must be $t(H, W)$. Hence, in other words, a graphon W is H-regular if the homomorphism density of H in W when one of the vertices of H is marked, is a constant independent of the value of the marking. Alternatively, H-regularity is equivalent to the condition that the U-statistic $\Delta_2(H,\, G_n)$ is first-order degenerate.

Remark 2.5. Recall that the *degree function* of a graphon W is defined as

$$d_W(x) := \int_0^1 W(x, y)\mathrm{d}y.$$

Note that when $H = K_2$ is the single edge,

$$t_1(x, K_2, W) = \mathbb{E}\left[W(U_1, U_2) \;\middle|\; U_1 = x\right] = \int_0^1 W(x, y)\mathrm{d}y = d_W(x).$$

Hence, the notion of K_2-regularity coincides with the standard notion of *degree regularity*, where the degree function $d_W(x) := \int_0^1 W(x, y)\mathrm{d}y$ is constant almost everywhere.

In the dense regime (where $\rho_n \asymp 1$), [5, 28] showed that H-regularity can lead $X(H, G_n)$ to have non-Gaussian fluctuations. The situation, however, is very different in the sparse regime. In this case, H-regularity implies that $\Delta_2(H, G_n)$ has asymptotically negligible contribution to the total variance of $X(H, G_n)$. Consequently, the asymptotic distribution of $X(H, G_n)$ is determined entirely by the fluctuations of $\Delta_1(H, G_n)$.

Theorem 2.6. *Suppose Assumption 2.3 holds and $W : [0,1]^2 \to \mathbb{R}_{\geqslant 0}$ is a H-regular graphon. Then,*

$$\frac{\Delta_2(H,\, G_n)}{\sqrt{\mathrm{Var}[\Delta(H,G_n)]}} \xrightarrow{L_2} 0 \quad \text{and} \quad \frac{\Delta_1(H,G_n)}{\sqrt{\mathrm{Var}[\Delta(H,G_n)]}} \xrightarrow{D} N(0,\,1), \tag{2.5}$$

Consequently,

$$Z(H,G_n) = \frac{\Delta(H,G_n)}{\sqrt{\mathrm{Var}[\Delta(H,G_n)]}} \xrightarrow{D} N(0,\,1), \tag{2.6}$$

for $Z(H,G_n)$ as defined in (2.2).

Next, we consider the case where W is not H-regular. In this case, the situation is more delicate. Specifically, an additional threshold on ρ_n emerges, which determines whether or not $\Delta_2(H,G_n)$ contributes to the fluctuations of $X(H,G_n)$. To describe this threshold we need the following definition:

Definition 2.7 (Strongly balanced graph)**.** [34, Equation (3.17)] For a fixed graph $H = (V(H), E(H))$ define

$$m_1(H) := \max_{F \subseteq H,\, |V(F)| \geqslant 2} \frac{|E(F)|}{|V(F)|-1}, \tag{2.7}$$

where the maximum is taken over all possible subgraphs of H with at least 2 vertices. If the maximum is attained at H, that is, $m_1(H) = \frac{|E(H)|}{|V(H)|-1}$, then H is said to be a *strongly balanced* graph. Moreover, if the maximum is attained uniquely at H, then H is called a *strictly strongly balanced graph.*

Cycles, trees and complete graphs are strongly balanced. Figure 2 (b) is an example of graph H which is not strongly balanced. In this case, for the subgraph F marked by the red dotted circle

$$\frac{|E(F)|}{|V(F)|-1} = \frac{3}{2} > \frac{4}{3} = \frac{|E(H)|}{|V(H)|-1}.$$

Remark 2.8. The notion of a *strongly balanced graph* was introduced by [53], where it is shown that $n\rho_n^{m_1(H)} \asymp 1$ is the threshold for the containment of a rooted copy of H in the Erdős-Rényi model $G(n,\rho_n)$. Recently, Maugis [44] showed that for the sparse graphon model the number of rooted copies of H (appropriately standardized) has Gaussian fluctuations, whenever $n\rho_n^{m_1(H)} \gg 1$. In Theorem 2.11 we prove that when W is not H-regular, then $n\rho_n^{m_1(H)} \asymp 1$ marks the threshold where the asymptotic behavior of $\Delta_1(H,G_n)$ and $\Delta_2(H,G_n)$ undergoes a phase transition:

- when $n\rho_n^{m_1(H)} \ll 1$, $\Delta_1(H,G_n)$ is asymptotically Gaussian and $\Delta_2(H,G_n)$ is asymptotically negligible;
- when $n\rho_n^{m_1(H)} \gg 1$, $\Delta_2(H,G_n)$ is asymptotically Gaussian and $\Delta_1(H,G_n)$ is asymptotically negligible; and
- when $n\rho_n^{m_1(H)} \asymp 1$, $(\Delta_1(H,G_n), \Delta_2(H,G_n))$ are asymptotically jointly Gaussian.

In the following proposition we collect a few basic implications of being strongly balanced. The proof is given in Appendix A.1.

Proposition 2.9. *For any graph H with $|E(H)| \geqslant 1$, $m(H) < m_1(H)$. Further, if H is strongly balanced, then H is strictly balanced.*

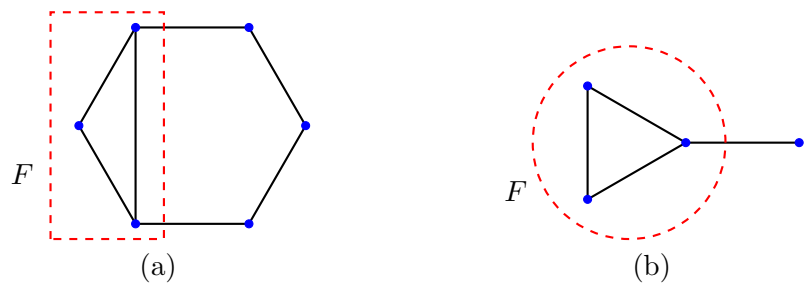

FIGURE 2. (a) A strictly balanced graph that is not strongly balanced; and (b) a balanced (but not strictly balanced) graph, which is not strongly balanced.

Remark 2.10. Proposition 2.9 shows that strongly balanced implies strictly balanced. The converse, however, is not true. Specifically, Figure 2 (a) shows a strictly balanced graph that is not strongly balanced. In this case, the subgraph F marked by the red dotted rectangle satisfies

$$\frac{|E(F)|}{|V(F)|-1} = \frac{3}{2} > \frac{7}{5} = \frac{|E(H)|}{|V(H)|-1}.$$

Further Figure 2 (b) is another example of graph that is balanced (but not strictly balanced), which is not strongly balanced.

The following result provides a complete characterization of the fluctuations of $\Delta_1(H, G_n)$ and $\Delta_2(H, G_n)$ and, as a result, that of $\Delta(H, G_n)$, when the graphon W is not H-regular.

Theorem 2.11. *Suppose Assumption 2.3 holds and the graphon W is not H-regular. Then the following hold:*

(1) *If $n\rho_n^{m_1(H)} \ll 1$, then*

$$\frac{\Delta_2(H,\, G_n)}{\sqrt{\mathrm{Var}[\Delta(H, G_n)]}} \xrightarrow{L_2} 0 \quad \textit{and} \quad \frac{\Delta_1(H,\, G_n)}{\sqrt{\mathrm{Var}[\Delta(H, G_n)]}} \xrightarrow{D} N(0,\, 1). \tag{2.8}$$

(2) *If $n\rho_n^{m_1(H)} \gg 1$, then*

$$\frac{\Delta_1(H,\, G_n)}{\sqrt{\mathrm{Var}[\Delta(H, G_n)]}} \xrightarrow{L_2} 0 \quad \textit{and} \quad \frac{\Delta_2(H,\, G_n)}{\sqrt{\mathrm{Var}[\Delta(H, G_n)]}} \xrightarrow{D} N(0,\, 1). \tag{2.9}$$

(3) *If $n\rho_n^{m_1(H)} \to c \in (0,\, \infty)$, then there exists a constant $\kappa \in (0,\, 1)$ (defined in (6.9)) such that*

$$\begin{pmatrix} \frac{\Delta_1(H, G_n)}{\sqrt{\mathrm{Var}[\Delta(H, G_n)]}} \\ \frac{\Delta_2(H, G_n)}{\sqrt{\mathrm{Var}[\Delta(H, G_n)]}} \end{pmatrix} \xrightarrow{D} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \kappa & 0 \\ 0 & 1-\kappa \end{pmatrix} \right). \tag{2.10}$$

Consequently, in all three cases,

$$Z(H, G_n) = \frac{\Delta(H, G_n)}{\sqrt{\mathrm{Var}[\Delta(H, G_n)]}} \xrightarrow{D} N(0,\, 1), \tag{2.11}$$

whenever $n^{-\frac{1}{m(H)}} \ll \rho_n \ll 1$.

In words, Theorem 2.11 shows that when W is not H-regular, then there are 3 possibilities depending on the strictly balanced coefficient $m_1(H)$:

- $n\rho_n^{m_1(H)} \ll 1$: In this case, $\Delta_2(H, G_n)$ has asymptotically negligible contribution to the total variance of $X(H, G_n)$ and the asymptotic distribution of $X(H, G_n)$ is determined entirely by the fluctuations of $\Delta_1(H, G_n)$.
- $n\rho_n^{m_1(H)} \gg 1$: In this case, $\Delta_1(H, G_n)$ has asymptotically negligible contribution to the total variance of $X(H, G_n)$ and the asymptotic distribution of $X(H, G_n)$ is determined entirely by the fluctuations of $\Delta_2(H, G_n)$.
- $n\rho_n^{m_1(H)} \to c \in (0, \infty)$: In this case, both $\Delta_1(H, G_n)$ and $\Delta_2(H, G_n)$ contribute to the fluctuations of $X(H, G_n)$. In particular, $(\Delta_1(H, G_n), \Delta_2(H, G_n))$ (after appropriate standardization) converges jointly independent normals with non-trivial variances.

Remark 2.12. Recall that Theorem 2.11 is under the standing assumption that $n\rho_n^{m(H)} \ll 1$. By Proposition 2.9 $m(H) < m_1(H)$ for every subgraph H with $|E(H)| \geqslant 1$. Hence, all the above three regimes in Theorem 2.11 are relevant.

Another common consequence Theorem 2.11 in all the three cases, is that $Z(H, G_n)$ is asymptotically normal. We know from Theorem 2.6 that the same is also true when W is H-regular (recall (2.6)). Hence, we have the following corollary:

Corollary 2.13. *Suppose Assumption 2.3 holds. Then the following hold:*

$$Z(H, G_n) = \frac{\Delta(H, G_n)}{\sqrt{\mathrm{Var}[\Delta(H, G_n)]}} \xrightarrow{D} N(0,\, 1),$$

where $Z(H, G_n)$ is as defined in (2.2).

This shows that $Z(H, G_n)$ is asymptotically normal (irrespective of whether or not W is H-regular), for any $\rho_n \ll 1$ above the containment threshold. In previous works, this result has been established for either certain types of graph H or in certain sub-regimes of sparsity above the containment threshold. Our result closes all the gaps in both these directions by showing that asymptotic normality of $Z(H, G_n)$ holds for all fixed graph H and for the full range of the sparsity parameter above the containment threshold. A detailed comparison with existing results is given in Section 2.4.2.

2.3. **A General Conditional CLT and Proof Outline.** The proofs of Theorem 2.6 and Theorem 2.11 rely on a result about the conditional fluctuations of $\Delta_1(H,\, G_n)$, which might be of independent interest. To state the result we need the following definition:

Definition 2.14. Given two real-valued random variables X and Z and a σ-algebra $\mathcal{F}$, the Wasserstein distance between $X \mid \mathcal{F}$ (the conditional law of X given the σ-algebra $\mathcal{F}$) and Z is defined as

$$d_{\mathrm{Wass}}(X \mid \mathcal{F},\, Z) := \sup_{g \in \mathcal{L}} \left| \mathbb{E}[g(X) \mid \mathcal{F}] - \mathbb{E}[g(Z)] \right|,$$

where $\mathcal{L}$ is the collection of 1-Lipschitz functions from $\mathbb{R} \to \mathbb{R}$.

We are now ready to state our result on the conditional convergence of $\Delta_1(H,\, G_n)$ (appropriately normalized). Recall that $\mathcal{F}(\boldsymbol{U}_n)$ is the sigma-algebra generated by the collection $\boldsymbol{U}_n = \{U_1, U_2, \ldots, U_n\}$.

Proposition 2.15. *Suppose Assumption 2.3 holds and* $\frac{\mathrm{Var}[\Delta_1(H, G_n)]}{\mathrm{Var}[\Delta(H, G_n)]} \to \kappa$ *for some constant* $\kappa > 0$. *Then, for* $Z \sim N(0, 1)$,

$$d_{\mathrm{Wass}}\left(\frac{\Delta_1(H,\, G_n)}{\sqrt{\mathrm{Var}[\Delta_1(H, G_n) \mid \mathcal{F}(\boldsymbol{U}_n)]}} \mid \mathcal{F}(\boldsymbol{U}_n), Z \right) \xrightarrow{P} 0. \tag{2.12}$$

The proof of Proposition 2.15 is given in Section 4. The proof uses on Stein's method based on dependency graphs, which bounds the distance to normality in the Wasserstein distance. Note that the Wasserstein distance in the LHS of (2.12) is itself random (since it is defined conditional on $\mathcal{F}(\boldsymbol{U}_n)$), hence the convergence in (2.12) is shown to hold in probability. An important feature of result in Proposition 2.15 is that it holds regardless of whether or not the graphon W is H-irregular, whenever $\Delta_1(H, G_n)$ has a non-trivial contribution to the total variance. Hence, Proposition 2.15 can used to prove the assertions of Theorem 2.6 and Theorem 2.11 as follows:

- *W is H-regular* (Theorem 2.6): This is proved in Section 5. The first step is a variance computation to show that
$$\mathrm{Var}[\Delta_2(H,\, G_n)] \ll \mathrm{Var}[\Delta(H, G_n)] \to 0.$$
Then, by invoking Proposition 2.15 the result in (2.5) follows.
- *W is is not H-regular* (Theorem 2.11): This is proved in Section 6. There are three cases:
  - $n\rho_n^{m_1(H)} \ll 1$: This is similar to the regular regime. A variance computation shows that
  $$\mathrm{Var}[\Delta_2(H,\, G_n)] \ll \mathrm{Var}[\Delta(H, G_n)] \to 0.$$
  Then, by invoking Proposition 2.15 the result in (2.8) follows.
  - $n\rho_n^{m_1(H)} \gg 1$: In this case, a variance computation shows that
  $$\mathrm{Var}[\Delta_1(H,\, G_n)] \ll \mathrm{Var}[\Delta(H, G_n)] \to 0.$$
  Hence, only $\Delta_2(H,\, G_n)$ contributes to the fluctuations of $\Delta(H, G_n)$. In this case, $\Delta_2(H,\, G_n)$ is U-statistic of order $|E(H)|$ which is not first-order degenerate W is is not H-irregular (since W is is not H-regular). Hence, invoking standard results about the asymptotic normality of U-statistics the result in (2.9) follows.
  - $n\rho_n^{m_1(H)} \to c \in (0, \infty)$: In this case, $\Delta_1(H,\, G_n)$ and $\Delta_2(H,\, G_n)$ both contribute to the fluctuations of $\Delta(H, G_n)$. In particular, by combining Proposition 2.15 (for $\Delta_1(H,\, G_n)$), the asymptotic normality of non-degenerate U-statistics (for $\Delta_2(H,\, G_n)$), and Lemma A.1 gives,
  $$\begin{pmatrix} \frac{\Delta_1(H, G_n)}{\sqrt{\mathrm{Var}[\Delta_1(H, G_n)]}} \\ \frac{\Delta_2(H, G_n)}{\sqrt{\mathrm{Var}[\Delta_2(H, G_n)]}} \end{pmatrix} \xrightarrow{D} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right).$$
  Then by a direct computation of the limit of $\frac{\mathrm{Var}[\Delta_1(H, G_n)]}{\mathrm{Var}[\Delta(H, G_n)]}$, the result in (2.11) follows.

2.4. **Prior Work.** In this section, we summarize prior work on the fluctuations of subgraph counts in inhomogeneous random graph models. For clarity, we organize the discussion into three subsections, focusing separately on the (homogeneous) Erdős-Rényi model, the sparse graphon model, and the dense graphon model.

2.4.1. *Erdős-Rényi Model.* In the Erdős-Rényi model $G(n, \rho_n)$ (which corresponds to $W \equiv 1$), it is well-known that $Z(H, G_n)$ is asymptotically normal if and only if $n\rho_n^{m(H)} \gg 1$ and $n^2(1-\rho_n) \gg 1$. This result was established independently by Ruciński [51] (using the method of moments), Barbour [3] (using Stein's method), Nowicki [46] (using U-statistics), and Janson [33] (using martingales), among others (see [34, Chapter 6] for a comprehensive treatment). Note that in the sparse regime $\rho_n \ll 1$, the condition $n^2(1-\rho_n) \gg 1$ is always satisfied, ensuring asymptotic

normality for all subgraph counts $X(H, G_n)$ whenever $n\rho_n^{m(H)} \gg 1$. Corollary 2.13 shows that the same result hold for the sparse W-random model as well.

2.4.2. *Sparse Graphon Model.* Central limit theorem for the subgraph counts in the sparse graphon model was first derived by Bickel et al. [8]. The sparsity conditions in their result depends on whether H is acyclic (that is, H does not contain a cycle) or whether it contains a cycle. Specifically, they showed the following results. Throughout, suppose $G_n \sim G(n, \rho_n, W)$.

- *H is acyclic*: In this case, [8, Theorem 1] shows that $X(H, G_n)$ (appropriately standardized) is asymptotically normal, whenever $n\rho_n \gg 1$.[2] Note that for any acyclic graph $m(H) = 1$. Hence, [8, Theorem 1] covers the full sparsity regime when H is acyclic.
- *H is acyclic*: In this case, [8, Theorem 1] shows that $X(H, G_n)$ is asymptotically normal, whenever $n\rho_n^{\frac{|V(H)|}{2}} \gg 1$. Recalling (2.1) note that

$$m(H) = \max_{F \subseteq H, |V(F)| \geqslant 1} \frac{|E(F)|}{|V(F)|} \leqslant \max_{F \subseteq H, |V(F)| \geqslant 1} \frac{\binom{|V(F)|}{2}}{|V(F)|} \leqslant \frac{|V(H)| - 1}{2} < \frac{|V(H)|}{2}.$$

  Hence, there is a gap between the sparsity regime covered by [8, Theorem 1] and the full sparsity regime ($n\rho_n^{m(H)} \gg 1$), when H contains a cycle.

Later, Zhang and Xia [56] strengthened above the results by proving Edgeworth expansions and Berry-Essen bounds, under stronger sparsity conditions. Very recently, using the method of cumulants Liu and Privault [40] derived normal approximation bounds in the Kolmogorov distance for $Z(H, G_n)$, when H is strongly balanced. Notably, the rate of convergence in [40, Corollary 7.8] depends on whether $n\rho_n^{m_1(H)} \ll 1$ or $n\rho_n^{m_1(H)} \gg 1$, which, interestingly, coincides with the threshold identified in Theorem 2.11, where the dominant contribution to the total variance shifts from $\Delta_1(H, G_n)$ and $\Delta_2(H, G_n)$. The results in [40], however, do not consider the case where H is not strongly balanced, or when $n\rho_n^{m_1(H)} \asymp 1$. In comparison, our results apply to all fixed subgraphs (not necessarily strongly balanced) and cover the entire sparsity regime, albeit without explicit rates of convergence.

2.4.3. *Dense Graphon Model.* Throughout this paper we have considered the sparse graphon model (where $\rho_n \ll 1$). In the dense regime (when $\rho_n = 1$), the asymptotic normality of $X(H, G_n)$ for all graphs H was established by Féray et al. [26] (see also [2, 23, 35, 57] for alternate proofs, extensions, and related results). However, the limiting normal distribution of $X(H, G_n)$ as derived in [26], can be degenerate depending on the structure of W. This degeneracy phenomenon was investigated by Hladkỳ et al. [28] when $H = K_r$ is the r-clique (the complete graph on r vertices), for general graphs H by Bhattacharya et al. [5], and for joint distributions by Chatterjee et al. [18]. Specifically, these results show that when W is not H-regular (recall Definition 2.4), then $X(H, G_n)$ is asymptotically Gaussian, with a normalization factor of $n^{|V(H)|-\frac{1}{2}}$. However, when W is H-regular, then the normalization factor becomes $n^{|V(H)|-1}$ and the limiting distribution of $X(H, G_n)$ has, in general, a Gaussian component and another independent (non-Gaussian) component which is a (possibly) infinite weighted sum of centered chi-squared random variables. This degeneracy phenomenon also appears in the subsequent work of Chatterjee and Huang [17] on the fluctuations of the largest eigenvalue. Recently, Huang et al. [32] established an invariance principle for $X(H, G_n)$ that encompasses higher-order degeneracies.

[2]Technically, Bickel et al. [8] considers the number of induced copies, but their proof can be easily adapted for $X(H, G_n)$.

3. Proof of Proposition 2.2

The proof of Proposition 2.2 is a standard application of the first and second moment methods. Specifically, it relies on the following lemma, which derives the asymptotic orders of the mean and variance of $X(H, G_n)$.

Lemma 3.1. *Suppose Assumption 2.3 holds. Then for $G_n \sim G(n, \rho_n, W)$,*

$$\mathbb{E}[X(H, G_n)] \asymp_{H,W} n^{|V(H)|}\rho_n^{|E(H)|}. \tag{3.1}$$

Moreover, for $\rho_n \ll 1$,

$$\mathrm{Var}[X(H, G_n)] \asymp_{H,W} \max_{F\subseteq H:|V(F)|\geqslant 1} \frac{n^{2|V(H)|}\rho_n^{2|E(H)|}}{n^{|V(F)|}\rho_n^{|E(F)|}}. \tag{3.2}$$

The proof of Lemma 3.1 is given in Section 3.1. Here, using this result, we complete the proof of Proposition 2.2. First, suppose $n\rho_n^{m(H)} \ll 1$. Then there are two cases:

- *H is balanced*: In this case, $m(H) = \frac{|E(H)|}{|V(H)|}$ and $n\rho_n^{m(H)} \ll 1$ implies $n^{|V(H)|}\rho_n^{|E(H)|} \ll 1$, that is, $\mathbb{E}[X(H, G_n)] \to 0$ (recall (3.1)). Therefore, by Markov's inequality,
$$\mathbb{P}(X(H, G_n) > 0) = \mathbb{P}(X(H, G_n) \geqslant 1) \leqslant \mathbb{E}[X(H, G_n)] \to 0.$$
- *H is unbalanced*: Let $F \subseteq H$ be a subgraph of H such that $m(H) = \frac{|E(F)|}{|V(F)|}$. Then $n\rho_n^{m(H)} \ll 1$ implies $n^{|V(F)|}\rho_n^{|E(F)|} \to 0$, that is, $\mathbb{E}[X(F, G_n)] \to 0$. This implies,
$$\mathbb{P}(X(H, G_n) > 0) \leqslant \mathbb{P}(X(F, G_n) > 0) \leqslant \mathbb{E}[X(F, G_n)] \to 0.$$

The above two cases combined establish the 0-statement in Proposition 2.2.

Next, suppose $n\rho_n^{m(H)} \gg 1$. Then by Lemma 3.1,

$$\frac{\mathrm{Var}[X(H, G_n)]}{(\mathbb{E}[X(H, G_n)])^2} \lesssim_{H,W} \max_{F\subseteq H,\, |V(F)|\geqslant 1} \frac{1}{n^{|V(F)|}\rho_n^{|E(F)|}}. \tag{3.3}$$

Note that the condition $n\rho_n^{m(H)} \gg 1$ implies $n^{|V(F)|}\rho_n^{|E(F)|} \gg 1$, for all $F \subseteq H$ such that $|V(F)| \geqslant 1$. Hence, the RHS of (3.3) tends to 0 whenever $n\rho_n^{m(H)} \gg 1$. This means,

$$\frac{X(H, G_n)}{\mathbb{E}[X(H, G_n)]} \xrightarrow{P} 1 \quad \Longrightarrow \quad X(H, G_n) \xrightarrow{P} \infty,$$

since $\mathbb{E}[X(H, G_n)] \to \infty$, whenever $n\rho_n^{m(H)} \gg 1$. This completes the proof of Proposition 2.2. □

3.1. Proof of Lemma 3.1.

The result in (3.1) is an immediate consequence of (1.3) $t(H, W) > 0$ (recall Assumption 2.3).

We now proceed with the proof of (3.2). To this end, let $\mathcal{H}_n = \{H_1, H_2, \ldots, H_M\}$ be the collection of all copies of H in the complete graph K_n. Here, $M = X(H, K_n)$ the number of copies of H in K_n. Now, define $I_s := \mathbf{1}\{H_s \subseteq G_n\}$, for $1 \leqslant s \leqslant M$. Then $\mathbb{E}[I_s] = \rho_n^{|E(H)|}t(H, W)$, for all $1 \leqslant s \leqslant M$. Furthermore,

$$\begin{aligned}
&\mathrm{Var}[X(H, G_n)] \\
&= \mathrm{Var}\left[\sum_{s=1}^{M} I_s\right] \\
&= \sum_{\substack{H_s,\, H_t\in\mathcal{H}_n \\ |V(H_s)\cap V(H_t)|\geqslant 1}} \mathrm{Cov}[I_s, I_t]
\end{aligned}$$

$$= \sum_{F \subseteq H : |V(F)| \geqslant 1} \sum_{\substack{H_s,\, H_t \in \mathcal{H}_n \\ H_s \cap H_t \simeq F}} \mathrm{Cov}[I_s,\, I_t]$$

$$= \sum_{F \subseteq H : |V(F)| \geqslant 1} \sum_{\substack{H_s,\, H_t \in \mathcal{H}_n \\ H_s \cap H_t \simeq F}} [\rho_n^{2|E(H)|-|E(F)|} t(H_s \cup H_t,\, W) - (\rho_n^{|E(H)|} t(H,\, W))^2]. \tag{3.4}$$

where $H_s \cup H_t = (V(H_s) \cup V(H_t),\, E(H_s) \cup E(H_t))$. The second last line follows since for any two copies $H_s,\, H_t \in \mathcal{H}_n$, the graph $H_s \cap H_t = (V(H_s) \cap V(H_t),\, E(H_s) \cap E(H_t))$ is isomorphic to some subgraph $F \subseteq H$.[3]

Now, assume that $\rho_n \ll 1$. Fix a subgraph $F \subseteq H$ such that $|V(F)| \geqslant 1$. By Lemma and fix copies $H_s,\, H_t \in \mathcal{H}_n$ so that $H_s \cap H_t$ is isomorphic to F. Then

$$\begin{aligned} \mathrm{Cov}[I_s,\, I_t] &\leqslant \mathbb{E}[I_s I_t] = \rho_n^{|E(H_s \cup H_t)|} t(H_s \cup H_t,\, W) \\ &\leqslant \rho_n^{2|E(H)|-|E(F)|} ||W||_\infty^{2|E(H)|-|E(F)|} \lesssim_{H,W} \rho_n^{2|E(H)|-|E(F)|}. \end{aligned} \tag{3.5}$$

Also, note that for a fixed subgraph $F \subseteq H$, the number of ordered pairs $(H_s,\, H_t)$ with $H_s \cap H_t \simeq F$ is $O_H(n^{2|V(H)|-|V(F)|})$. Hence, (3.4) can be expressed as

$$\begin{aligned} \mathrm{Var}[X(H, G_n)] &\lesssim_{H,W} \sum_{F \subseteq H : |V(F)| \geqslant 1} n^{2|V(H)|-|V(F)|} \rho_n^{2|E(H)|-|E(F)|} \\ &\lesssim_{H,W} \max_{F \subseteq H : |V(F)| \geqslant 1} \frac{n^{2|V(H)|} \rho_n^{2|E(H)|}}{n^{|V(F)|} \rho_n^{|E(F)|}}. \end{aligned} \tag{3.6}$$

This proves the upper bound in (3.2)

To show the lower bound, we need the following lemma. The proof is given in Appendix A.2.

Lemma 3.2. *Let H be a subgraph with $|E(H)| \geqslant 1$ and $t(H,\, W) > 0$. Let $F \subseteq H$ be a subgraph of H with $|V(F)| \geqslant 1$. Then there exist copies $H_s,\, H_t \in \mathcal{H}_n$ of H in K_n such that $H_s \cap H_t \simeq F$ and $t(H_s \cup H_t,\, W) \geqslant (t(H,\, W))^2 > 0$. Moreover, the number of such pairs $H_s,\, H_t$ is $O_H(n^{2|V(H)|-|V(F)|})$.*

Fix a subgraph $F \subseteq H$ such that $|V(F)| \geqslant 1$. By Lemma 3.2, there are $O_H(n^{2|V(H)|-|V(F)|})$ pairs $H_s,\, H_t \in \mathcal{H}_n$ such that $H_s \cap H_t \simeq F$ and $t(H_s \cup H_t,\, W) \geqslant (t(H,\, W))^2 > 0$. This means, when $|E(F)| \geqslant 1$,

$$\begin{aligned} \mathrm{Cov}[I_s,\, I_t] &= \mathbb{E}[I_s I_t] - (\rho_n^{|E(H)|} t(H,\, W))^2 \\ &= \rho_n^{2|E(H)|-|E(F)|} t(H_s \cup H_t,\, W) - \rho_n^{2|E(H)|} (t(H,\, W))^2 \gtrsim_{H,W} \rho_n^{2|E(H)|-|E(F)|}, \end{aligned}$$

since $\rho_n \ll 1$. Also, when $|E(F)| = 0$, by Lemma 3.2, we have $t(H_s \cup H_t,\, W) \geqslant (t(H,\, W))^2$. and, hence,

$$\begin{aligned} \mathrm{Cov}[I_s,\, I_t] &= \mathbb{E}[I_s I_t] - (\rho_n^{|E(H)|} t(H,\, W))^2 \\ &= \rho_n^{2|E(H)|} (t(H_s \cup H_t,\, W) - (t(H,\, W))^2) \gtrsim_{H,W} \rho_n^{2|E(H)|-|E(F)|}. \end{aligned}$$

Thus, (3.4) leads to

$$\begin{aligned} \mathrm{Var}[X(H, G_n)] &\gtrsim_{H,W} \sum_{F \subseteq H : |V(F)| \geqslant 1} n^{2|V(H)|-|V(F)|} \rho_n^{2|E(H)|-|E(F)|} \\ &\gtrsim_{H,W} \max_{F \subseteq H : |V(F)| \geqslant 1} \frac{n^{2|V(H)|} \rho_n^{2|E(H)|}}{n^{|V(F)|} \rho_n^{|E(F)|}}. \end{aligned} \tag{3.7}$$

[3]For two graphs G_1 and G_2, the notation $G_1 \simeq G_2$ will mean that G_1 and G_2 are isomorphic.

Combining (3.6) and (3.7), the result in (3.2) follows. □

4. Proof of Proposition 2.15

The proof of Proposition 2.15 will use Stein's method based on dependency graphs. We begin by defining the notion of a conditional dependency graph, given a sigma-algebra.

Definition 4.1 (Conditional dependency graph)**.** Let $\{X_v\}_{v\in V}$ be a family of random variables (defined on some common probability space) indexed by a finite set V and $\mathcal{F}$ be a σ-algebra. Then a graph $\mathcal{G}$ with vertex set V is said to be a conditional dependency graph for the collection $\{X_v\}$ given $\mathcal{F}$ if the following holds: For any two subsets of vertices $A, B \subseteq V$ such that there is no edge from in $\mathcal{G}$ from any vertex in A to any vertex in B, then the colelctions $\{X_v\}_{v\in A}$ and $\{X_v\}_{v\in B}$ are conditionally independent given $\mathcal{F}$.

For a dependency graph $\mathcal{G}$ and $u \in V$, denote by $\bar{N}_{\mathcal{G}}(u)$ the set neighbors of u in $\mathcal{G}$ and u itself. Also, denote $\bar{N}_{\mathcal{G}}(u,\, v) := \bar{N}_{\mathcal{G}}(u) \cup \bar{N}_{\mathcal{G}}(v)$. We will use the following version of Stein's method based on dependency graph.

Proposition 4.2 ([34, Theorem 6.33])**.** *Suppose $\{X_v\}_{v\in V}$ be a family of random variables with conditional dependency graph $\mathcal{G}$ given $\mathcal{F}$. Assume $\mathbb{E}(X_v|\mathcal{F}) = 0$ almost surely, for all $v \in V$. Let*

$$W = \frac{1}{\sigma(\mathcal{F})} \sum_{v\in V} X_v,$$

where $(\sigma(\mathcal{F}))^2 = \mathrm{Var}[\sum_{v\in V} X_v \mid \mathcal{F}]$ is the conditional variance. Assume, for all $u,\, v \in V$,

$$\sum_{v\in V} \mathbb{E}[|X_v| \mid \mathcal{F}] \leqslant R(\mathcal{F}) \quad \text{and} \quad \sum_{w\in \bar{N}_{\mathcal{G}}(u,\, v)} \mathbb{E}[|X_w| \mid X_u,\, X_v,\, \mathcal{F}] \leqslant Q(\mathcal{F}), \tag{4.1}$$

almost surely. Then, for $Z \sim N(0, 1)$,

$$d_{\mathrm{Wass}}(W \mid \mathcal{F},\, Z) \lesssim \frac{R(\mathcal{F})(Q(\mathcal{F}))^2}{(\sigma(\mathcal{F}))^3}$$

almost surely.

Remark 4.3. Technically, [34, Theorem 6.33] establishes the above result in the unconditional setting. The conditional result stated above follows by repeating the proof verbatim (with expectations replaced by conditional expectations).

With the above preparations we can now proceed with the proof of Proposition 2.15. To begin with, let $\mathcal{H}_n = \{H_1,\, H_2,\, \ldots,\, H_M\}$ be the collection of copies of all H in the complete graph K_n. For $1 \leqslant s \leqslant M$, define $I_s := \mathbf{1}\{H_j \subseteq G_n\}$ and

$$X_s = I_s - \mathbb{E}(I_s \mid \mathcal{F}(\boldsymbol{U}_n)) = I_s - \rho_n^{|E(H)|} \prod_{(a,\, b)\in E(H_s)} W(U_a,\, U_b).$$

Clearly, $\mathbb{E}[X_s \mid \mathcal{F}] = 0$, for all $1 \leqslant s \leqslant M$. Recalling the definition from $\Delta_1(H,\, G_n)$ (1.5), note that

$$\Delta_1(H,\, G_n) = \sum_{s=1}^{M} X_s.$$

Construct a conditional dependency graph $\mathcal{G}$ of the collection of random variables $\{X_s : H_s \in \mathcal{H}_n\}$ given $\sigma(\boldsymbol{U}_n)$ on the vertex set $\{1,\, 2,\, \ldots,\, M\}$ as follows: Connect the edge $(s,\, t)$, for $1 \leqslant s \leqslant M$ in $\mathcal{G}_n$ if and only if $|E(H_s) \cap E(H_t)| \geqslant 1$.

With this dependency graph, we now bound the terms appearing in Proposition 4.2. We begin the first term in (4.1).

Lemma 4.4. *We have*

$$\sum_{s=1}^{M} \mathbb{E}[|X_s| \mid \mathcal{F}(\boldsymbol{U}_n)] \lesssim_{H,W} \mathbb{E}[X(H, G_n)], \tag{4.2}$$

almost surely.

Proof. For $1 \leqslant s \leqslant M$,

$$\begin{aligned}\mathbb{E}[|X_s| \mid \mathcal{F}(\boldsymbol{U}_n)] &= 2\mathbb{E}[I_s \mid \mathcal{F}(\boldsymbol{U}_n)](1 - \mathbb{E}[I_s \mid \mathcal{F}(\boldsymbol{U}_n)]) \\ &= 2\rho_n^{|E(H)|} \prod_{(a,b)\in E(H_s)} W(U_a, U_b) \left(1 - \rho_n^{|E(H)|} \prod_{(a,b)\in E(H_s)} W(U_a, U_b)\right) \\ &\lesssim_{H,W} \rho_n^{|E(H)|} ||W||_\infty^{|E(H)|} \lesssim_{H,W} \rho_n^{|E(H)|} \end{aligned} \tag{4.3}$$

almost surely. Further, with $M = |\mathcal{H}_n| = \binom{n}{|V(H)|} \frac{|V(H)|!}{|\mathrm{Aut}(H)|} \lesssim_{H,W} n^{|V(H)|}$,

$$\sum_{s=1}^{M} \mathbb{E}[|X_s| \mid \mathcal{F}(\boldsymbol{U}_n)] \lesssim_{H,W} n^{|V(H)|} \rho_n^{|E(H)|} \asymp_{H,W} \mathbb{E}[X(H, G_n)] \tag{4.4}$$

where the last step uses (3.1). Thus, Lemma 4.4 follows. □

Next, we consider the second term in (4.1).

Lemma 4.5. *For all* $1 \leqslant s < t \leqslant M$

$$\sum_{w\in \bar{N}_{\mathcal{G}}(s,t)} \mathbb{E}[|X_w| \mid X_s, X_t, \mathcal{F}(\boldsymbol{U}_n)] \lesssim_{H,W} \frac{\mathbb{E}[X(H, G_n)]}{\Phi_H}, \tag{4.5}$$

almost surely, where

$$\Phi_H := \min\{\mathbb{E}[X(F, G_n)] : F \subseteq H, |V(F)| \geqslant 1\} \asymp \min_{F\subseteq H, |V(F)|\geqslant 1} n^{|V(F)|} \rho_n^{|E(F)|}.$$

Proof. Fix $H_s, H_t \in \mathcal{H}_n$, for $1 \leqslant s < t \leqslant M$. For every $H_w \in \mathcal{H}_n$, define $H_{w,\{s,t\}} = H_w \cap (H_s \cup H_t)$. Now, observe that $H_{w,\{s,t\}}$ is isomorphic to some subgraph $F \subseteq H$, since H_w is a copy of H. Then, almost surely,

$$\begin{aligned}\mathbb{E}[|X_w| \mid X_s, X_t, \mathcal{F}(\boldsymbol{U}_n)] &\leqslant \mathbb{E}[I_w \mid I_s, I_t, \mathcal{F}(\boldsymbol{U}_n)] + \rho_n^{|E(H)|} \prod_{(a,b)\in E(H_w)} W(U_a, U_b) \\ &= \rho_n^{|E(H)|-|E(F)|} \prod_{(a,b)\in E(H_w)\setminus E(H_{w,\{s,t\}})} W(U_a, U_b) \\ &\qquad + \rho_n^{|E(H)|} \prod_{(a,b)\in E(H_w)} W(U_a, U_b) \\ &\leqslant \rho_n^{|E(H)|-|E(F)|} ||W||_\infty^{|E(H)|-|E(F)|} + \rho_n^{|E(H)|} ||W||_\infty^{|E(H)|} \\ &\lesssim_{H,W} \rho_n^{|E(H)|-|E(F)|}, \end{aligned} \tag{4.6}$$

since $\rho_n \ll 1$. If we fix a subgraph $F \subseteq H$, then the number of choices of $w \in \{1, 2, \ldots, M\}$ such that $H_{w,\{s,t\}}$ is isomorphic to F is $O_H(n^{|V(H)|-|V(F)|})$. Furthermore, note that $w \in \bar{N}_{\mathcal{G}}(s, t)$ if and only if $|E(H_{w,\{s,t\}})| > 0$. Hence, due to (4.6), (4.5) can be bounded as:

$$\sum_{w\in \bar{N}_{\mathcal{G}}(s,t)} \mathbb{E}[|X_w| \mid X_s, X_t, \mathcal{F}(\boldsymbol{U}_n)] \lesssim_{H,W} \sum_{F\subseteq H: |E(F)|\geqslant 1} n^{|V(H)|-|V(F)|} \rho_n^{|E(H)|-|E(F)|}$$

$$\lesssim_{H,W} \max_{F\subseteq H:|V(F)|\geqslant 1} n^{|V(H)|-|V(F)|}\rho_n^{|E(H)|-|E(F)|}$$
$$\asymp_{H,W} \frac{\mathbb{E}[X(H,\, G_n)]}{\Phi_H},$$

almost surely due to (3.1). Thus Lemma 4.4 follows. □

Finally, to apply Proposition 4.2 we need a lower bound on the conditional variance $\sigma(\mathcal{F}(\boldsymbol{U}_n))$ where

$$(\sigma(\mathcal{F}(\boldsymbol{U}_n)))^2 = \mathrm{Var}[\Delta_1(H,\, G_n)|\mathcal{F}(\boldsymbol{U}_n)] = \sum_{\substack{H_s,\, H_t\in\mathcal{H}_n\\ |E(H_s)\cap E(H_t)|\geqslant 1}} \mathrm{Cov}[I_s,\, I_t \mid \mathcal{F}(\boldsymbol{U}_n)], \tag{4.7}$$

since the conditional covariance is 0 when $|E(H_s)\cap E(H_t)| = 0$.

Lemma 4.6. *Suppose* $\frac{\mathrm{Var}[\Delta_1(H,G_n)]}{\mathrm{Var}[\Delta(H,G_n)]} \to \kappa$ *for some constant* $\kappa > 0$. *Then*

$$(\sigma(\mathcal{F}(\boldsymbol{U}_n)))^2 \gtrsim_{H,W} \frac{(\mathbb{E}[X(H,G_n)])^2}{\Phi_H}, \tag{4.8}$$

almost surely.

Proof. Note that if $|E(H_s)\cap E(H_t)| \geqslant 1$, then

$$\mathrm{Cov}[I_s,\, I_t \mid \mathcal{F}(\boldsymbol{U}_n)] = \rho_n^{|E(H_s)\cup E(H_t)|} \prod_{(a,\,b)\in E(H_s)\cup E(H_t)} W(U_a,U_b) - \rho_n^{2|E(H)|}A_n(s,t)$$

where

$$A_n(s,t) := \prod_{(a,\,b)\in E(H_s)} W(U_a,U_b) \prod_{(c,d)\in E(H_t)} W(U_c,\, U_d).$$

Note that $H_s\cap H_t = (V(H_s)\cap V(H_t),\, E(H_s)\cap E(H_t))$ is isomorphic to a subgraph F of H. This means, from (4.7),

$$(\sigma(\mathcal{F}(\boldsymbol{U}_n)))^2 = \sum_{\substack{H_s,\, H_t\in\mathcal{H}_n\\ |E(H_s)\cap E(H_t)|\geqslant 1}} \mathrm{Cov}[I_s,\, I_t \mid \sigma(\boldsymbol{U}_n)]$$

$$= \sum_{\substack{F\subseteq H\\ |E(F)|\geqslant 1}} \left\{ \rho_n^{2|E(H)|-|E(F)|} \sum_{\substack{H_s,\, H_t\in\mathcal{H}_n\\ H_s\cap H_t\simeq F}} \prod_{(a,\,b)\in E(H_s)\cup E(H_t)} W(U_a,U_b) - \rho_n^{2|E(H)|} \sum_{\substack{H_s,\, H_t\in\mathcal{H}_n\\ H_s\cap H_t\simeq F}} A_n(s,t) \right\}$$

$$= \sum_{\substack{F\subseteq H\\ |E(F)|\geqslant 1}} \left\{ \rho_n^{2|E(H)|-|E(F)|} Z_n(F) - \rho_n^{2|E(H)|} A_n(F) \right\}, \tag{4.9}$$

almost surely, where

$$Z_n(F) := \sum_{\substack{H_s,\, H_t\in\mathcal{H}_n\\ H_s\cap H_t\simeq F}} \prod_{(a,\,b)\in E(H_s)\cup E(H_t)} W(U_a,U_b) \text{ and } A_n(F) := \sum_{\substack{H_s,\, H_t\in\mathcal{H}_n\\ H_s\cap H_t\simeq F}} A_n(s,t). \tag{4.10}$$

Observe that $Z_n(F)$ and $A_n(F)$ are U-statistics with kernels of order $2|V(H)| - |V(F)|$. Hence, from the strong law for U-statistics (see [16, Theorem 6.3.1]) it follows that $\frac{Z_n(F)}{\mathbb{E}[Z_n(F)]} \to 1$ and $\frac{A_n(F)}{\mathbb{E}[A_n(F)]} \to 1$ almost surely. Thus,

$$\frac{\sigma(\mathcal{F}(\boldsymbol{U}_n))^2}{\mathrm{Var}[\Delta_1(H,\, G_n)]} = \frac{\sigma(\mathcal{F}(\boldsymbol{U}_n))^2}{\mathbb{E}[\mathrm{Var}[\Delta_1(H,\, G_n) \mid \mathcal{F}(\boldsymbol{U}_n)]]} \qquad (\text{since } \mathbb{E}[\Delta_1(H,\, G_n))|\mathcal{F}(\boldsymbol{U}_n)] = 0)$$

$$= \frac{\sum_{F\subseteq H:|E(F)|\geqslant 1}\left\{\rho_n^{2|E(H)|-|E(F)|}Z_n(F) - \rho_n^{2|E(H)|}A_n(F)\right\}}{\sum_{F\subseteq H:|E(F)|\geqslant 1}\left\{\rho_n^{2|E(H)|-|E(F)|}\mathbb{E}[Z_n(F)] - \rho_n^{2|E(H)|}\mathbb{E}[A_n(F)]\right\}}$$
$$\stackrel{a.s.}{\to} 1, \tag{4.11}$$

since there are finitely many terms in the sums in both numerator and denominator and for every fixed F, $\frac{Z_n(F)}{\mathbb{E}[Z_n(F)]} \to 1$ and $\frac{A_n(F)}{\mathbb{E}[A_n(F)]} \to 1$, almost surely. Since we assumed that $\frac{\mathrm{Var}[\Delta_1(H,G_n)]}{\mathrm{Var}[\Delta(H,G_n)]} \to \kappa$ for some constant $\kappa > 0$, using (3.2) gives,

$$\mathrm{Var}[\Delta_1(H,\, G_n)] \asymp_{H,W} \mathrm{Var}[\Delta(H, G_n)] \asymp_{H,W} \max_{F\subseteq H:|V(F)|\geqslant 1} n^{2|V(H)|-|V(F)|}\rho_n^{2|E(H)|-|E(F)|}$$
$$\asymp \frac{(\mathbb{E}[X(H, G_n)])^2}{\Phi_H} \tag{4.12}$$

due to Lemma 3.1. Combining (4.11) and (4.12), the result in (4.8) follows. □

Now, applying in the bounds (4.2), (4.5), and (4.8), in Proposition 4.2 gives

$$d_{\mathrm{Wass}}\left((\sigma(\mathcal{F}(\boldsymbol{U}_n)))^{-1}\Delta_1(H,\, G_n)|\mathcal{F}(\boldsymbol{U}_n),\, Z\right) \lesssim_{H,W} \Phi_H^{-\frac{1}{2}},$$

where $Z \sim N(0,\, 1)$. Note that $\Phi_H \gg 1$ when $n\rho_n^{m(H)} \gg 1$. This completes the proof of Proposition 2.15. □

5. Proof of Theorem 2.6

Recalling the definition of $\Delta_2(H,\, G_n)$ from (1.5) note that

$$\begin{aligned}
&\Delta_2(H,\, G_n)\\
&= \mathbb{E}[\Delta(H,\, G_n|\boldsymbol{U}_n)] - \mathbb{E}[\Delta(H, G_n)]\\
&= \sum_{1\leqslant i_1<\ldots<i_{|V(H)|}\leqslant n}\ \sum_{H'\in\mathscr{G}_H(\{i_1,\ldots,i_{|V(H)|}\})}\left(\prod_{(i_s,\,i_t)\in E(H')}\rho_n W(U_{i_s},\, U_{i_t}) - \rho_n^{|E(H)|}t(H,\, W)\right)\\
&= \rho_n^{|E(H)|}\sum_{1\leqslant i_1<\ldots<i_{|V(H)|}\leqslant n} h(U_{i_1},\, U_{i_2},\, \ldots,\, U_{i_{|V(H)|}}),
\end{aligned}$$

where

$$h(x_{i_1},\, x_{i_2},\, \ldots,\, x_{i_{|V(H)|}}) = \sum_{H'\in\mathscr{G}_H(\{i_1,\ldots,i_{|V(H)|}\})}\left(\prod_{(i_s,\,i_t)\in E(H')} W(x_{i_s},\, x_{i_t}) - t(H,\, W)\right).$$

Hence, defining

$$T(H,\, G_n) = \frac{1}{\binom{n}{|V(H)|}}\sum_{1\leqslant i_1<\ldots<i_{|V(H)|}\leqslant n} h(U_{i_1},\, U_{i_2},\, \ldots,\, U_{i_{|V(H)|}}), \tag{5.1}$$

we get

$$\Delta_2(H,\, G_n) = \binom{n}{|V(H)|}\rho_n^{|E(H)|}T(H,\, G_n) \tag{5.2}$$

Note that $T(H,\, G_n)$ is a centered U-statistics of order $|V(H)|$ with kernel h. In the next lemma we show that the variance of $T(H, G_n)$ scales as $\frac{1}{n}$ when the graphon W is H-regular. The proof is deferred to Appendix A.3.

Lemma 5.1. *$T(H, G_n)$ is first-order degenerate if and only if the graphon W is H-regular. Consequently, when W is H-regular,* $\mathrm{Var}[T(H, G_n)] \lesssim_{H,W} \frac{1}{n^2}$.

Theorem 2.6 assumes that W is H-regular. Now recalling (5.2), and using Lemma 5.1,

$$\mathrm{Var}[\Delta_2(H, G_n)] \lesssim_{H,W} n^{2|V(H)|-2}\rho_n^{2|E(H)|}.$$

This means, recalling (3.2),

$$\frac{\mathrm{Var}[\Delta_2(H, G_n)]}{\mathrm{Var}[\Delta(H, G_n)]} \lesssim_{H,W} \min_{F\subseteq H:|V(F)|\geqslant 1} n^{|V(F)|-2}\rho_n^{|E(F)|} \to 0. \tag{5.3}$$

This proves the first assertion in (2.5). Next, using (2.3), $\frac{\mathrm{Var}[\Delta_1(H, G_n)]}{\mathrm{Var}[\Delta(H, G_n)]} = 1 - \frac{\mathrm{Var}[\Delta_2(H, G_n)]}{\mathrm{Var}[\Delta(H, G_n)]} \to 1$. This means, from (4.11), $\frac{\mathrm{Var}[\Delta_1(H, G_n)|\mathcal{F}(\boldsymbol{U}_n)]}{\mathrm{Var}[\Delta(H, G_n)]} \to 1$. Hence, applying Proposition 2.15 and Slutsky's theorem it follows that

$$\frac{\Delta_1(H, G_n)}{\sqrt{\mathrm{Var}[\Delta(H, G_n)]}} \xrightarrow{D} N(0, 1),$$

unconditionally. This proves the second assertion in (2.5). The result is (2.6) is an immediate consequence of (1.5) and (2.5). □

6. Proof of Theorem 2.11

In this section, we prove Theorem 2.11, which considers the case where W is H-regular. We begin by showing that $\Delta_2(H, G_n)$ (respectively, $\Delta_1(H, G_n)$) has negligible contribution to the total variance when $n\rho_n^{m_1(H)} \ll 1$ (respectively, $n\rho_n^{m_1(H)} \gg 1$).

Proposition 6.1. *Suppose W is not H-regular. Then the following holds:*

1. *If $n\rho_n^{m_1(H)} \ll 1$, then* $\mathrm{Var}[\Delta_2(H, G_n)] \ll \mathrm{Var}[\Delta(H, G_n)]$.
2. *If $n\rho_n^{m_1(H)} \gg 1$, then* $\mathrm{Var}[\Delta_1(H, G_n)] \ll \mathrm{Var}[\Delta(H, G_n)]$.

Proof of Proposition 6.1. Recall the definition of $T(H, G_n)$ from (5.1). Since W is not H-regular, by Lemma 5.1 $T(H, G_n)$ is an U-statistic of order $|V(H)|$ which is non-degenerate. Hence, by [37, Section 1.3 Theorem 3], $\mathrm{Var}[T(H, G_n)] \lesssim_{H,W} \frac{1}{n}$, and recalling (5.2),

$$\mathrm{Var}[\Delta_2(H, G_n)] \lesssim_{H,W} n^{2|V(H)|-1}\rho_n^{2|E(H)|}.$$

Also, recalling (3.2),

$$\mathrm{Var}[\Delta(H, G_n)] \asymp_{H,W} \max_{F\subseteq H:|V(F)|\geqslant 1} n^{2|V(H)|-|V(F)|}\rho_n^{2|E(H)|-|E(F)|}. \tag{6.1}$$

We claim that when $n\rho_n^{m_1(H)} \ll 1$, the maximum in (6.1) is obtained for some subgraph $F \subseteq H$ with $|V(F)| \geqslant 2$. Indeed, when $n\rho_n^{m_1(H)} \ll 1$, there exists subgraph $F \subseteq H$ with $|V(F)| \geqslant 2$ such that $n^{|V(F)|-1}\rho_n^{|E(F)|} \ll 1$ or equivalently, $n^{2|V(H)|-1}\rho_n^{2|E(H)|} \ll n^{2|V(H)|-|V(F)|}\rho_n^{2|E(H)|-|E(F)|}$. Thus,

$$\frac{\mathrm{Var}[\Delta_2(H, G_n)]}{\mathrm{Var}[\Delta(H, G_n)]} \lesssim_{H,W} \min_{F\subseteq H,\,|V(F)|\geqslant 2} n^{|V(F)|-1}\rho_n^{|E(F)|} \to 0$$

since $n\rho_n^{m_1(H)} \ll 1$. This proves Proposition 6.1 (1).

Next, for the second part, from [37, Section 1.3, Theorem 3] we know that

$$\lim_{n\to\infty} n\mathrm{Var}[T(H, G_n)] = |V(H)|^2\xi_1,$$

where ξ_1 is defined in Appendix A.3. Note that $\xi_1 > 0$, as W is not H-regular. Thus, recalling (5.2),

$$\mathrm{Var}[\Delta_2(H,\, G_n)] = \frac{|V(H)|^2}{n}\binom{n}{|V(H)|}^2 \rho_n^{2|E(H)|}\xi_1 + o(n^{2|V(H)|-1}\rho_n^{2|E(H)|}). \tag{6.2}$$

Now, continuing from (3.4), we have

$$\mathrm{Var}[\Delta(H,\, G_n)] = \sum_{F\subseteq H:|V(F)|=1}\ \sum_{\substack{H_s,\, H_t\in\mathcal{H}_n\\ H_s\cap H_t\simeq F}} \mathrm{Cov}[I_s,\, I_t] + \sum_{F\subseteq H:|V(F)|\geqslant 2}\ \sum_{\substack{H_s,\, H_t\in\mathcal{H}_n\\ H_s\cap H_t\simeq F}} \mathrm{Cov}[I_s,\, I_t], \tag{6.3}$$

where F is the subgraph of H isomorphic to $H_s \cap H_t = (V(H_s)\cap V(H_t),\, E(H_s)\cap E(H_t))$. Note that for a fixed subgraph $F \subseteq H$, the number of pairs $(H_s,\, H_t)$ having $H_s \cap H_t$ isomorphic to F is $O_H(n^{2|V(H)|-|V(F)|})$. Hence, recalling (3.5), the second term in (6.3) can be bounded as

$$\begin{aligned}\sum_{F\subseteq H:|V(F)|\geqslant 2}\ \sum_{\substack{H_s,\, H_t\in\mathcal{H}_n\\ H_s\cap H_t\simeq F}} \mathrm{Cov}[I_s,\, I_t] &\lesssim_{H,\,W} \sum_{F\subseteq H:|V(F)|\geqslant 2} n^{2|V(H)|-|V(F)|}\rho_n^{2|E(H)|-|E(F)|}\\ &= o(n^{2|V(H)|-1}\rho_n^{2|E(H)|}),\end{aligned}$$

since $n\rho_n^{m_1(H)} \gg 1$ is equivalent to the condition that for all subgraph $F \subseteq H$ with $|V(F)| \geqslant 2$, we have $n^{2|V(H)|-|V(F)|}\rho_n^{2|E(H)|-|E(F)|} \ll n^{2|V(H)|-1}\rho_n^{2|E(H)|}$. To compute the first term in (6.3), first choose the vertex sets of the two copies of H sharing a single vertex. This is equivalent to finding the number of ordered pairs of subsets $(A_1,\, A_2)$ of $[n]$ such that $|A_1 \cap A_2| = 1$. Observe that the number of such ordered pairs is $|V(H)|\binom{n}{|V(H)|}\binom{n-|V(H)|}{|V(H)|-1}$. Once we have fixed the vertex sets of the two copies, we can trace the steps of the proof of Lemma 5.1 to obtain

$$\mathrm{Var}[\Delta(H, G_n)] = |V(H)|\binom{n}{|V(H)|}\binom{n-|V(H)|}{|V(H)|-1}\rho_n^{2|E(H)|}\xi_1 + o(n^{2|V(H)|-1}\rho_n^{2|E(H)|}). \tag{6.4}$$

from (6.3). Taking the ratio of (6.2) and (6.4) it follows that $\frac{\mathrm{Var}[\Delta_2(H,G_n)]}{\mathrm{Var}[\Delta(H,G_n)]} \to 1$. Hence, recalling (2.3) the result in Proposition 6.1 (2) follows. □

Using the above result, we now complete the proof of Theorem 2.11. To begin with suppose, $n\rho_n^{m_1(H)} \ll 1$. The first assertion in (2.8) follows from Proposition 6.1 (1). This means, using (2.3), $\frac{\mathrm{Var}[\Delta_1(H,G_n)]}{\mathrm{Var}[\Delta(H,G_n)]} = 1 - \frac{\mathrm{Var}[\Delta_2(H,G_n)]}{\mathrm{Var}[\Delta(H,G_n)]} \to 1$. Hence, from (4.11), $\frac{\mathrm{Var}[\Delta_1(H,G_n)|\mathcal{F}(\boldsymbol{U}_n)]}{\mathrm{Var}[\Delta(H,G_n)]} \to 1$ almost surely. Hence, applying Proposition 2.15 and Slutsky's theorem it follows that

$$\frac{\Delta_1(H,\, G_n)}{\sqrt{\mathrm{Var}[\Delta(H, G_n)]}} \xrightarrow{D} N(0,\, 1),$$

unconditionally. This proves the second assertion in (2.8).

Next, suppose $n\rho_n^{m_1(H)} \gg 1$. The first assertion in (2.9) follows from Proposition 6.1 (2). Also, since W is not H-regular, by Lemma 5.1 $\Delta_2(H,\, G_n)$ is a scaled U-statistic of order $|V(H)|$ which is non-degenerate. Hence, by [37, Chapter 3, Theorem 1],

$$\frac{\Delta_2(H,\, G_n)}{\sqrt{\mathrm{Var}[\Delta_2(H,\, G_n)]}} \xrightarrow{D} N(0, 1). \tag{6.5}$$

Then by Slutsky's lemma

$$\frac{\Delta_2(H,\, G_n)}{\sqrt{\mathrm{Var}[\Delta(H,\, G_n)]}} \xrightarrow{D} N(0, 1).$$

This proves the second assertion in (2.9).

Next, consider the case $n\rho_n^{m_1(H)} \to c$, for some $c \in (0, \infty)$. We have, from (6.2),

$$\mathrm{Var}[\Delta_2(H, G_n)] = n^{2|V(H)|-1}\rho_n^{2|E(H)|}\frac{|V(H)|^2}{(|V(H)|!)^2}\xi_1 + o(n^{2|V(H)|-1}\rho_n^{2|E(H)|}).$$

Also, we know from (4.9),

$$\begin{aligned}\mathrm{Var}[\Delta_1(H, G_n)] &= \mathbb{E}[\mathrm{Var}[\Delta_1(H, G_n)|\mathcal{F}(\boldsymbol{U}_n)]] \\ &= \sum_{\substack{F\subseteq H\\ |E(F)|\geqslant 1}} \left\{\rho_n^{2|E(H)|-|E(F)|}\mathbb{E}[Z_n(F)] - \rho_n^{2|E(H)|}\mathbb{E}[A_n(F)]\right\}\end{aligned}$$

where $Z_n(F)$ and $A_n(F)$ are defined in (4.10). Note that under Assumption 2.3 and Lemma 3.2,

$$\mathbb{E}[\rho_n^{2|E(H)|-|E(F)|}Z_n(F))] \asymp_{H,W} n^{2|V(H)|-|V(F)|}\rho_n^{2|E(H)|-|E(F)|}$$

and $\rho_n^{2|E(H)|}\mathbb{E}[A_n(F)] = o(n^{2|V(H)|-|V(F)|}\rho_n^{2|E(H)|-|E(F)|})$. Hence,

$$\mathrm{Var}[\Delta_1(H, G_n)] = \sum_{\substack{F\subseteq H\\ |E(F)|\geqslant 1}} \left\{\rho_n^{2|E(H)|-|E(F)|}\mathbb{E}[Z_n(F)] + o(n^{2|V(H)|-|V(F)|}\rho_n^{2|E(H)|-|E(F)|})\right\} \quad (6.6)$$

When $n\rho_n^{m_1(H)} \to c \in (0, \infty)$, for any $F \subseteq H$ with $|V(F)| \geqslant 2$, we have $n^{|V(F)|-1}\rho_n^{|E(F)|} \gtrsim_{H,W} 1$, or, equivalently, $n^{2|V(H)|-|V(F)|}\rho_n^{2|E(H)|-|E(F)|} \lesssim_{H,W} n^{2|V(H)|-1}\rho_n^{2|E(H)|}$. Thus, continuing from (6.6),

$$\mathrm{Var}[\Delta_1(H, G_n)] = \sum_{\substack{F\subseteq H\\ |E(F)|\geqslant 1}} \rho_n^{2|E(H)|-|E(F)|}\mathbb{E}[Z_n(F)] + o(n^{2|V(H)|-1}\rho_n^{2|E(H)|}).$$

To find the dominating term in above sum, define the following set:

$$\mathscr{S}(H) := \left\{F \subseteq H : |E(F)| \geqslant 1 \text{ and } m_1(H) = \frac{|E(F)|}{|V(F)|-1}\right\}.$$

This is the collection of subgraphs F of H where the maximum in (2.7) is attained. (Note that if H is strictly strongly balanced $\mathscr{S}(H) = \{H\}$.)

Lemma 6.2. *Suppose $n\rho_n^{m_1(H)} \to c \in (0, \infty)$. If $F \subseteq H$ is a subgraph with $|E(F)| \geqslant 1$ and $F \notin \mathscr{S}(H)$, then $\mathbb{E}[\rho_n^{2|E(H)|-|E(F)|}Z_n(F)] \ll n^{2|V(H)|-1}\rho_n^{2|E(H)|}$.*

Proof. Fix a subgraph $F \subseteq H$ with $|E(F)| \geqslant 1$ and $F \notin \mathscr{S}(H)$. Then $\frac{|E(F)|}{|V(F)|-1} < m_1(H)$. Since $n\rho_n^{m_1(H)} \to c \in (0, \infty)$, we must have $n^{|V(F)|-1}\rho_n^{|E(F)|} \gg 1$, or equivalently,

$$n^{2|V(H)|-|V(F)|}\rho_n^{2|E(H)|-|E(F)|} \ll n^{2|V(H)|-1}\rho_n^{2|E(H)|}. \quad (6.7)$$

There are $O_H(n^{2|V(H)|-|V(F)|})$ copies H_s, $H_t \in \mathcal{H}_n$ of H such that $H_s \cup H_t \simeq F$. Thus, according to the definition of $Z_n(F)$ in (4.10), we have

$$\begin{aligned}\mathbb{E}[\rho_n^{2|E(H)|-|E(F)|}Z_n(F)] &\lesssim_{H,W} n^{2|V(H)|-|V(F)|}\rho_n^{2|E(H)|-|E(F)|}||W||_\infty^{2|E(H)|-|E(F)|} \\ &\lesssim_{H,W} n^{2|V(H)|-|V(F)|}\rho_n^{2|E(H)|-|E(F)|} \ll n^{2|V(H)|-1}\rho_n^{2|E(H)|},\end{aligned}$$

where the last line follows from (6.7). This completes the proof of the lemma. □

The above lemma implies,

$$\frac{\mathrm{Var}[\Delta_2(H, G_n)]}{\mathrm{Var}[\Delta(H, G_n)]}$$

$$= \frac{\mathrm{Var}[\Delta_2(H,\, G_n)]}{\mathrm{Var}[\Delta_1(H, G_n)] + \mathrm{Var}[\Delta_2(H, G_n)]} \tag{6.8}$$
$$= \frac{n^{2|V(H)|-1}\rho_n^{2|E(H)|}\frac{|V(H)|^2}{(|V(H)|!)^2}\xi_1 + o(n^{2|V(H)|-1}\rho_n^{2|E(H)|})}{n^{2|V(H)|-1}\rho_n^{2|E(H)|}\frac{|V(H)|^2}{(|V(H)|!)^2}\xi_1 + \sum_{F\in\mathscr{S}(H)}\mathbb{E}[\rho_n^{2|E(H)|-|E(F)|}Z_n(F)] + o(n^{2|V(H)|-1}\rho_n^{2|E(H)|})}.$$

Hence, to compute the limit of (6.8) we need to consider graphs which are in $\mathscr{S}(H)$. Fix graphs H and $F \subseteq H$, and define

$$\mathscr{T}(H,\, F) := \{R : H_s \cup H_t \simeq R, \text{ where } H_s,\, H_t \simeq H \text{ with } H_s \cap H_t \simeq F\}.$$

Also, define $\eta(H,\, F,\, R)$ to be the number of ways two copies of H can be joined on the vertex set $\{1,\, 2,\, \ldots,\, 2|V(H)| - |V(F)|\}$ such that their intersection is isomorphic to F and their union is isomorphic to R. Formally, given $F \subseteq H$ and $R \in \mathscr{T}(H,\, F)$, we have

$$\eta(H,\, F,\, R) = |\{(H_1,\, H_2) : H_1 \simeq H_2 \simeq H,\, H_1 \cap H_2 \simeq F,\, H_1 \cup H_2 \simeq R\}|.$$

Lemma 6.3. *Suppose $n\rho_n^{m_1(H)} \to c \in (0,\, \infty)$. If $F \in \mathscr{S}(H)$, then*

$$\frac{\mathbb{E}[\rho_n^{2|E(H)|-|E(F)|}Z_n(F)]}{n^{2|V(H)|-1}\rho_n^{2|E(H)|}} = \frac{1}{c^{|V(F)|-1}(2|V(H)| - |V(F)|)!}\sum_{R\in\mathscr{T}(H,\,F)}\eta(H,\, F,\, R)t(R,\, W) + o(1).$$

Proof. As $F \in \mathscr{S}(H)$, we have $n^{|V(F)|-1}\rho_n^{|E(F)|} \to c^{|V(F)|-1}$. Then,

$$\begin{aligned}
&\frac{\mathbb{E}(\rho_n^{2|E(H)|-|E(F)|}Z_n(F))}{n^{2|V(H)|-1}\rho_n^{2|E(H)|}} \\
&= \frac{\rho_n^{2|E(H)|-|E(F)|}}{n^{2|V(H)|-1}\rho_n^{2|E(H)|}}\binom{n}{2|V(H)| - |V(F)|}\sum_{R\in\mathscr{T}(H,\,F)}\eta(H,\, F,\, R)t(R,\, W) \\
&\sim \frac{n^{2|V(H)|-|V(F)|}\rho_n^{2|E(H)|-|E(F)|}}{n^{2|V(H)|-1}\rho_n^{2|E(H)|}(2|V(H)| - |V(F)|)!}\sum_{R\in\mathscr{T}(H,\,F)}\eta(H,\, F,\, R)t(R,\, W) \\
&\sim \frac{1}{c^{|V(F)|-1}(2|V(H)| - |V(F)|)!}\sum_{R\in\mathscr{T}(H,\,F)}\eta(H,\, F,\, R)t(R,\, W).
\end{aligned}$$

This completes the proof of Lemma 6.3. □

Applying Lemma 6.2 and Lemma 6.3 to (6.8) gives,

$$\kappa := 1 - \frac{\frac{|V(H)|^2}{(|V(H)|!)^2}\xi_1}{\frac{|V(H)|^2}{(|V(H)|!)^2}\xi_1 + \sum_{F\in\mathscr{S}(H)}\frac{1}{c^{|V(F)|-1}(2|V(H)|-|V(F)|)!}\sum_{R\in\mathscr{T}(H,\,F)}\eta(H,\, F,\, R)t(R,\, W)}. \tag{6.9}$$

Clearly, $\kappa \in (0, 1)$. This shows (recall (2.3)),

$$\frac{\mathrm{Var}[\Delta_1(H,\, G_n)]}{\mathrm{Var}[\Delta(H, G_n)]} \to \kappa \quad \text{and} \quad \frac{\mathrm{Var}[\Delta_2(H,\, G_n)]}{\mathrm{Var}[\Delta(H, G_n)]} \to 1 - \kappa. \tag{6.10}$$

Now, using Proposition 2.15, (6.5) (which also holds when $n\rho_n^{m_1(H)} \to c \in (0,\, \infty)$ by the same arguments), and Lemma A.1, we have

$$\begin{pmatrix} \frac{\Delta_1(H, G_n)}{\sqrt{\mathrm{Var}[\Delta_1(H,G_n)]}} \\ \frac{\Delta_2(H, G_n)}{\sqrt{\mathrm{Var}[\Delta_2(H,G_n)]}} \end{pmatrix} \xrightarrow{D} N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right).$$

Then applying Slutsky's theorem with (6.10) proves the result in (2.10). □

Remark 6.4. Note that when H is strictly strongly balanced, that is, $\mathscr{S}(H) = \{H\}$, then expression of κ in (6.9) simplifies to:

$$\begin{aligned}\kappa &= 1 - \frac{\frac{|V(H)|^2}{(|V(H)|!)^2}\xi_1}{\frac{|V(H)|^2}{(|V(H)|!)^2}\xi_1 + \frac{1}{c^{|V(H)|-1}|V(H)|!}|\mathscr{G}_H|t(H,\, W)} \\ &= \frac{t(H,\, W)/|\mathrm{Aut}(H)|}{t(H,\, W)/|\mathrm{Aut}(H)| + c^{|V(H)|-1}\xi_1/((|V(H)|-1)!)^2},\end{aligned}$$

recalling (1.2). Here, $|\mathscr{G}_H|$ is the number of copies of H in the complete graph $K_{|V(H)|}$ on $|V(H)|$ vertices, and ξ_1 is defined in Appendix A. This shows that under $n\rho_n^{m_1(H)} \to c \in (0,\, \infty)$,

$$\frac{\mathrm{Var}[\Delta_1(H,\, G_n)]}{\mathrm{Var}[\Delta(H, G_n)]} \to \kappa \quad \text{and} \quad \frac{\mathrm{Var}[\Delta_2(H,\, G_n)]}{\mathrm{Var}[\Delta(H, G_n)]} \to 1-\kappa,$$

where $\kappa \in (0,\, 1)$ is a constant depending on c, H and W.

7. Conclusions

In this paper, we show that $X(H, G_n)$, with $G_n \sim G(n, \rho_n, W)$, is asymptotically normal whenever $n\rho_n^{m(H)} \gg 1$, that is, when the sparsity level is above the containment threshold for H. We do not consider the critical regime $n\rho_n^{m(H)} \asymp 1$. For the classical Erdős–Rényi model, it is known that at the threshold, $X(H, G_n)$ converges to a Poisson distribution when H is strictly balanced [10]. If H is unbalanced, the problem reduces to a balanced subgraph via an appropriate normalization [52, Section 4]. When H is balanced but not strictly balanced, the limiting distribution is more intricate and depends on the structure and multiplicity of subgraphs maximizing $m(H)$, with no universal closed form [11, 34].

Recently, there has been growing interest in deriving analogous results for inhomogeneous random graph models. In this direction, Coulson et al. [21] derived Poisson approximation results for $X(H, G_n)$, when G_n is sampled from a stochastic block model and H is strictly balanced (see also [22]). Recently, Liu and Privault [39] obtained an analogous Poisson approximation result in the random connection model. Extending these results to the sparse graphon model, as well as addressing the case where H is not strictly balanced, are interesting directions for future research.

Acknowledgement. BBB and SC were supported in part by NSF CAREER grant DMS 2046393 and a Sloan Research Fellowship.

References

[1] A. Athreya, D. E. Fishkind, M. Tang, C. E. Priebe, Y. Park, J. T. Vogelstein, K. Levin, V. Lyzinski, Y. Qin, and D. L. Sussman. Statistical inference on random dot product graphs: a survey. *Journal of Machine Learning Research*, 18(226):1–92, 2018.

[2] M. Austern and P. Orbanz. Limit theorems for distributions invariant under groups of transformations. *The Annals of Statistics*, 50(4):1960–1991, 2022.

[3] A. D. Barbour, M. Karoński, and A. Ruciński. A central limit theorem for decomposable random variables with applications to random graphs. *Journal of Combinatorial Theory, Series B*, 47(2): 125–145, 1989.

[4] A. Basak and S. Mukherjee. Universality of the mean-field for the potts model. *Probability Theory and Related Fields*, 168(3):557–600, 2017.

[5] B. B. Bhattacharya, A. Chatterjee, and S. Janson. Fluctuations of subgraph counts in graphon based random graphs. *Combinatorics, Probability and Computing*, 32(3):428–464, 2023.

[6] S. Bhattacharyya and P. J. Bickel. Subsampling bootstrap of count features of networks. *The Annals of Statistics*, 43(6):2384 – 2411, 2015.

[7] P. J. Bickel and A. Chen. A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073, 2009.

[8] P. J. Bickel, A. Chen, and E. Levina. The method of moments and degree distributions for network models. *The Annals of Statistics*, 39(5):2280–2301, 2011.

[9] M. Boguná and R. Pastor-Satorras. Class of correlated random networks with hidden variables. *Physical Review E*, 68(3):036112, 2003.

[10] B. Bollobás. Threshold functions for small subgraphs. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 90(2), pages 197–206. Cambridge University Press, 1981.

[11] B. Bollobás and J. C. Wierman. Subgraph counts and containment probabilities of balanced and unbalanced subgraphs in a large random graph a. *Annals of the New York Academy of Sciences*, 576 (1):63–70, 1989.

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

[13] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6): 1801–1851, 2008.

[14] C. Borgs, J. Chayes, and L. Lovász. Moments of two-variable functions and the uniqueness of graph limits. *Geometric and Functional Analysis*, 19:1597–1619, 2010.

[15] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs II. multiway cuts and statistical physics. *Annals of Mathematics*, pages 151–219, 2012.

[16] A. Bose, A. Chakrabarty, and R. S. Hazra. *A Little Book of Martingales*. Springer, 2024.

[17] A. Chatterjee and J. Huang. Fluctuation of the largest eigenvalue of a kernel matrix with application in graphon-based random graphs. *arXiv:2401.01866*, 2024.

[18] A. Chatterjee, S. Dan, and B. B. Bhattacharya. Higher-order graphon theory: Fluctuations, degeneracies, and inference. *arXiv:2404.13822*, 2024.

[19] S. Chatterjee and P. Diaconis. Estimating and understanding exponential random graph models. *The Annals of Statistics*, 41(5):2428–2461, 2013.

[20] S. Chatterjee and S. S. Varadhan. The large deviation principle for the Erdős-Rényi random graph. *European Journal of Combinatorics*, 32(7):1000–1017, 2011.

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

[22] M. Coulson, R. E. Gaunt, and G. Reinert. Compound poisson approximation of subgraph counts in stochastic block models with multiple edges. *Advances in Applied Probability*, 50(3):759–782, 2018.

[23] J.-F. Delmas, J.-S. Dhersin, and M. Sciauveau. Asymptotic for the cumulative distribution function of the degrees and homomorphism densities for random graphs sampled from a graphon. *Random Structures & Algorithms*, 58(1):94–149, 2021.

[24] P. Diaconis and D. Freedman. On the statistics of vision: the Julesz conjecture. *Journal of Mathematical Psychology*, 24(2):112–138, 1981.

[25] P. Erdős, A. Rényi, et al. On the evolution of random graphs. *Publ. math. inst. hung. acad. sci*, 5 (1):17–60, 1960.

[26] V. Féray, P.-L. Méliot, and A. Nikeghbali. Graphons, permutons and the Thoma simplex: three mod-Gaussian moduli spaces. *Proceedings of the London Mathematical Society*, 121(4):876–926, 2020.

[27] A. Green and C. R. Shalizi. Bootstrapping exchangeable random graphs. *Electronic Journal of Statistics*, 16(1):1058–1095, 2022.

[28] J. Hladkỳ, C. Pelekis, and M. Šileikis. A limit theorem for small cliques in inhomogeneous random graphs. *Journal of Graph Theory*, 97(4):578–599, 2021.

[29] P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of The American Statistical Association*, 97(460):1090–1098, 2002.

[30] P. W. Holland and S. Leinhardt. Transitivity in structural models of small groups. *Comparative Group Studies*, 2(2):107–124, 1971.

[31] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.
[32] K. H. Huang, M. Austern, and P. Orbanz. Gaussian universality for approximately polynomial functions of high-dimensional data. *arXiv:2403.10711*, 2024.
[33] S. Janson. A functional limit theorem for random graphs with applications to subgraph count statistics. *Random Structures & Algorithms*, 1(1):15–37, 1990.
[34] S. Janson, T. Łuczak, and A. Rucinski. *Random Graphs*. John Wiley & Sons, 2011.
[35] G. Kaur and A. Röllin. Higher-order fluctuations in dense random graph models. *Electronic Journal of Probability*, 26:1–36, 2021.
[36] O. Klopp, A. B. Tsybakov, and N. Verzelen. Oracle inequalities for network models and sparse graphon estimation. *Annals of Statistics*, 45(1):316–354, 2017.
[37] A. J. Lee. *U-statistics: Theory and Practice*. Routledge, 2019.
[38] Q. Lin, R. Lunde, and P. Sarkar. Trading off accuracy for speedup: Multiplier bootstraps for subgraph counts. *arXiv:2009.06170v5*, 2022.
[39] Q. Liu and N. Privault. Normal to poisson phase transition for subgraph counting in the random-connection model. *Electronic Journal of Probability*, 30:1–32, 2025.
[40] Q. Liu and N. Privault. Gaussian fluctuations of generalized u-statistics and subgraph counting in the binomial random-connection model. *Stochastic Processes and their Applications*, 192:104825, 2026.
[41] L. Lovász. *Large Networks and Graph Limits*, volume 60. American Mathematical Soc., 2012.
[42] L. Lovász and B. Szegedy. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957, 2006.
[43] R. Lunde and P. Sarkar. Subsampling sparse graphons under minimal assumptions. *Biometrika*, 110 (1):15–32, 2023.
[44] P.-A. Maugis. Central limit theorems for local network statistics. *Biometrika*, 111(3):743–754, 2024.
[45] P.-A. Maugis, S. C. Olhede, C. E. Priebe, and P. J. Wolfe. Testing for equivalence of network distribution using subgraph counts. *Journal of Computational and Graphical Statistics*, 29(3):455–465, 2020.
[46] K. Nowicki. Asymptotic normality of graph statistics. *Journal of statistical planning and inference*, 21(2):209–222, 1989.
[47] M. Penrose. *Random Geometric Graphs*, volume 5. OUP Oxford, 2003.
[48] N. Pržulj. Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2): e177–e183, 2007.
[49] P. Rubin-Delanchy, J. Cape, M. Tang, and C. E. Priebe. A statistical interpretation of spectral embedding: The generalised random dot product graph. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(4):1446–1473, 2022.
[50] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.
[51] A. Ruciński. When are small subgraphs of a random graph normally distributed? *Probability Theory and Related Fields*, 78(1):1–10, 1988.
[52] A. Rucinski. Small subgraphs of random graphs–a survey. In *Random graphs*, volume 87, pages 283–303, 1990.
[53] A. Ruciński and A. Vince. Strongly balanced graphs and random graphs. *Journal of graph theory*, 10(2):251–264, 1986.
[54] M. Shao, D. Xia, Y. Zhang, Q. Wu, and S. Chen. Higher-order accurate two-sample network inference and network hashing. *arXiv:2208.07573*, 2022.
[55] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *nature*, 393(6684): 440–442, 1998.
[56] Y. Zhang and D. Xia. Edgeworth expansions for network moments. *The Annals of Statistics*, 50(2): 726–753, 2022.
[57] Z.-S. Zhang. Berry–Esseen bounds for generalized U-statistics. *Electronic Journal of Probability*, 27: 1–36, 2022.

Appendix A. Proofs of Technical Lemmas

In this section we collect the proofs of some technical lemmas. We begin with the proof of Proposition 2.9 in Appendix A.1. Lemma 3.2 is proved in Appendix A.2. Then, in Appendix A.3 we prove Lemma 5.1. Finally, in Appendix A.4 we prove a result about establishing joint convergence from marginal convergence.

A.1. **Proof of Proposition 2.9.** Since $|E(H)| \geqslant 1$, the maximum in the definition of $m(H)$ is not attained at a single isolated vertex. Thus,

$$m(H) := \max_{F \subseteq H,\, |V(F)| \geqslant 1} \frac{|E(F)|}{|V(F)|} = \max_{F \subseteq H,\, |V(F)| \geqslant 2} \frac{|E(F)|}{|V(F)|} < \max_{F \subseteq H,\, |V(F)| \geqslant 2} \frac{|E(F)|}{|V(F)| - 1} = m_1(H),$$

which proves the first part of the proposition.

Next, suppose H is strongly balanced, that is, for all subgraphs $F = (V(F), E(F))$ of H, $\frac{|E(F)|}{|V(F)|-1} \leqslant \frac{|E(H)|}{|V(H)|-1}$. This means, for all F such that $|V(F)| < |V(H)|$,

$$\begin{aligned}\frac{|E(F)|}{|V(F)|} = \frac{|E(F)|(|V(F)|-1)}{(|V(F)|-1)|V(F)|} \leqslant \frac{|E(H)|(|V(F)|-1)}{(|V(H)|-1)|V(F)|} &\leqslant \frac{|E(H)|}{|V(H)|}\left(\frac{|V(H)|(|V(F)|-1)}{|V(F)|(|V(H)|-1)}\right) \\ &< \frac{|E(H)|}{|V(H)|}.\end{aligned}$$

Moreover, if $|V(F)| = |V(H)|$, then for any subgraph F which is not the whole graph H, $\frac{|E(F)|}{|V(F)|} < \frac{|E(H)|}{|V(H)|}$ since $|E(H)| \geqslant 1$. This shows strongly balanced graphs are strictly balanced. □

A.2. **Proof of Lemma 3.2.** Fix a subgraph $F \subseteq H$ with $|V(F)| \geqslant 1$. Consider two copies H_s, $H_t \in \mathcal{H}_n$ of H in K_n such that

$$\begin{aligned}V(H_s) &= \{i_1, i_2, \ldots, i_{|V(F)|}, i_{|V(F)|+1}, \ldots, i_{|V(H)|}\} \\ V(H_t) &= \{i_1, i_2, \ldots, i_{|V(F)|}, i_{|V(H)|+1}, \ldots, i_{2|V(H)|-|V(F)|}\},\end{aligned}$$

where the first $|V(F)|$ vertices are common. The edge sets of H_s, H_t are determined by maps $\phi_s : V(H) \to V(H_s)$ and $\phi_t : V(H) \to V(H_t)$, such that $(a, b) \in E(H)$ if and only if $(\phi_s(a), \phi_s(b)) \in E(H_s)$ and $(\phi_t(a), \phi_t(b)) \in E(H_t)$. In particular, we consider the following maps

$$\phi_s(a) = i_a,\ 1 \leqslant a \leqslant |V(H)|, \quad \text{and} \quad \phi_t(a) = \begin{cases} i_a & \text{if } 1 \leqslant a \leqslant |V(F)|, \\ i_{a+|V(H)|-|V(F)|} & \text{if } |V(F)|+1 \leqslant a \leqslant |V(H)|. \end{cases}$$

Then it is easy to see that the first $|V(F)|$ vertices of H_s and H_t constitute a copy of F, that is, $H_s \cap H_t \simeq F$, and the remaining part of the copy of H is extended in a similar fashion for both H_s and H_t. Then

$$\begin{aligned}t(H_s \cup H_t, W) &= \int_{[0,1]^{|V(H_s)\cup V(H_t)|}} \prod_{(a,b)\in E(H_s)\cup E(H_t)} W(U_a, U_b) \prod_{a \in V(H_s)\cup V(H_t)} \mathrm{d}x_a \\ &= \int_{[0,1]^{|V(F)|}} \frac{t^2_{V(F)}(x_1, x_2, \ldots, x_{|V(F)|}; H, W)}{\prod_{(a,b)\in E(F)} W(x_a, x_b)} \prod_{a \in V(F)} \mathrm{d}x_a,\end{aligned}$$

where

$$t_{V(F)}(x_1, x_2, \ldots, x_{|V(F)|}; H, W) = \mathbb{E}\left[\prod_{(a,b)\in E(H)} W(U_a, U_b) \mid U_{a_j} = x_j, \text{ for } 1 \leqslant j \leqslant |V(F)|\right],$$

is the $|V(F)|$-point conditional homomorphism density of H in W (see [5, Definition 2.1]). Then, by the Cauchy-Schwarz inequality,

$$t(H_s \cup H_t,\, W) \geqslant \frac{1}{||W||_\infty^{|E(F)|}} \left(\int_{[0,\,1]^{|V(F)|}} t_{V(F)}(x_1,\, x_2,\, \ldots,\, x_{|V(F)|};\, H,\, W) \prod_{a \in V(F)} \mathrm{d}x_a \right)^2$$
$$\geqslant t(H,\, W)^2 > 0.$$

Furthermore, note that to create pairs H_s, H_t as above, we may first choose the common $|V(F)|$ vertices in at most $\binom{n}{|V(F)|}$ ways. The rest of the vertices of H_s and H_t can be chosen in at most

$$\binom{n - |V(F)|}{|V(H)| - |V(F)|} \binom{n - |V(H)|}{|V(H)| - |V(F)|}$$

ways. Hence, the number of such pairs is $O_H(n^{2|V(H)| - |V(F)|})$. This completes the proof of the lemma. □

A.3. **Proof of Lemma 5.1.** We begin by recalling the expression of $T(H, G_n)$ from (5.1) and noting that $T(H,\, G_n)$ is degenerate if and only if

$$\xi_1 := \mathrm{Cov}[h(U_{i_1},\, U_{i_2},\, \ldots,\, U_{i_{|V(H)|}}),\, h(U_{i'_1},\, U_{i'_2},\, \ldots,\, U_{i'_{|V(H)|}})] = 0,$$

with $|\{i_1,\, i_2,\, \ldots,\, i_{|V(H)|}\} \cap \{i'_1,\, i'_2,\, \ldots,\, i'_{|V(H)|}\}| = 1$. We will show that $\xi_1 = 0$ if and only if the graphon W is H-regular. To prove this, observe that

$$\begin{aligned}
\xi_1 &= \mathrm{Cov}(h(U_1,\, U_2,\, \ldots,\, U_{|V(H)|}),\, h(U_{|V(H)|},\, U_{|V(H)|+1},\, \ldots,\, U_{2|V(H)|-1})) \\
&= \mathbb{E}(h(U_1,\, U_2,\, \ldots,\, U_{|V(H)|}) h(U_{|V(H)|},\, U_{|V(H)|+1},\, \ldots,\, U_{2|V(H)|-1})) \\
&= \sum_{\substack{H' \in \mathscr{G}_H(\{1,2,\ldots,|V(H)|\}) \\ H'' \in \mathscr{G}_H(\{|V(H)|,|V(H)|+1,\ldots,2|V(H)|-1\})}} \left[\mathbb{E} \left(\prod_{\substack{(i,j) \in E(H') \\ (k,l) \in E(H'')}} W(U_i,\, U_j) W(U_k,\, U_l) \right) - t(H,\, W)^2 \right].
\end{aligned}$$

To further simplify the expression we recall [5, Definition 2.7]: For $a, b \in \{1, 2, \ldots, |V(H)|\}$ the (a, b)-vertex join of two copies of H is the graph obtained by identifying the a-th vertex of first copy of H with the b-th vertex of the second copy of H. The resulting graph is denoted by $H \bigoplus_{a,b} H$. Now, applying [5, Lemma 5.2] gives,

$$\xi_1 = \frac{|\mathscr{G}_H|^2}{|V(H)|^2} \left[\sum_{1 \leqslant a,\, b \leqslant |V(H)|} t\left(H \bigoplus_{a,\, b} H,\, W \right) - |V(H)|^2 (t(H,\, W))^2 \right].$$

Finally, invoking [5, Lemma 5.4] we conclude that $\xi_1 = 0$ if and only if W is H-regular. Thus, when W is H-regular, $T(H, G_n)$ is a first order degenerate U-statistics. Hence by [37, Section 1.3, Theorem 3], $\mathrm{Var}[T(H,\, G_n)] \lesssim_{H,\, W} \frac{1}{n^2}$. This completes the proof of Lemma 5.1. □

A.4. **Joint Convergence from Marginal Convergence.** In this section we formulate a general result about bivariate convergence given the convergence of one of the marginals and the conditional convergence of the other marginal. This has been used in the proof of (2.10).

Lemma A.1. *Suppose $\{(X_n, Y_n)\}_{n \geqslant 1}$ be a collection of random variables and $\{\mathcal{F}_n\}_{n \geqslant 1}$ a collection of sigma algebras such Y_n is $\mathcal{F}_n$-measurable. Assume the following holds:*

- $d_{\mathrm{Wass}}(X_n \mid \mathcal{F}_n,\, Z) \xrightarrow{P} 0$, *where* $Z \sim N(0,\, 1)$.
- $Y_n \xrightarrow{D} N(0,\, 1)$.

Then

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{D} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right).$$

Proof. Note that it is enough to show,

$$\mathbb{E}[e^{isX_n + itY_n}] \to e^{-\frac{s^2+t^2}{2}},$$

for all $s,\, t \in (-1,\, 1)$. Notice that Y_n is $\mathcal{F}_n$-measurable. This implies,

$$\begin{aligned} \mathbb{E}[e^{isX_n + itY_n}] &= \mathbb{E}[e^{itY_n}\mathbb{E}[e^{isX_n} \mid \mathcal{F}_n]] \\ &= \mathbb{E}[e^{itY_n}(\mathbb{E}[e^{isX_n} \mid \mathcal{F}_n] - e^{-\frac{s^2}{2}})] + e^{-\frac{s^2}{2}}\mathbb{E}[e^{itY_n}]. \end{aligned} \tag{A.1}$$

Since $Y_n \xrightarrow{D} N(0,\, 1)$, we have $\mathbb{E}[e^{itY_n}] \to e^{-\frac{t^2}{2}}$, for all $t \in (-1,\, 1)$. Thus, it is enough to show that the first term in (A.1) converges to 0. To this end, note that,

$$\begin{aligned} \left|\mathbb{E}[e^{itY_n}\left(\mathbb{E}[e^{isX_n} \mid \mathcal{F}_n] - e^{-\frac{s^2}{2}}\right)\right| &\leqslant \mathbb{E}\left[\left|e^{itY_n}\right|\left|\mathbb{E}[e^{isX_n} \mid \mathcal{F}_n] - e^{-\frac{s^2}{2}}\right|\right] \\ &\leqslant \mathbb{E}\left[\left|\mathbb{E}[e^{isX_n} \mid \mathcal{F}_n] - \mathbb{E}[e^{isZ}]\right|\right]. \end{aligned} \tag{A.2}$$

To complete the proof note that,

$$\begin{aligned} &\left|\mathbb{E}[e^{isX_n} \mid \mathcal{F}_n] - \mathbb{E}[e^{isZ}]\right| \\ &\leqslant |\mathbb{E}[\cos(sX_n) \mid \mathcal{F}_n] - \mathbb{E}[\cos(sZ)]| + |\mathbb{E}[\sin(sX_n) \mid \mathcal{F}_n] - \mathbb{E}[\sin(sZ)]| \\ &\leqslant 2d_{\text{Wass}}(X_n \mid \mathcal{F}_n,\, Z) \xrightarrow{P} 0. \end{aligned}$$

The above convergence follows from Definition 2.14, since $x \mapsto \cos(sx)$ and $x \mapsto \sin(sx)$ are both 1-Lipschitz functions for $|s| < 1$. Also, as the conditional characteristic functions $\mathbb{E}[e^{isX_n} \mid \mathcal{F}_n]$ is uniformly bounded by 1, we have L_1 convergence, that is,

$$\mathbb{E}\left[\left|\mathbb{E}[e^{isX_n} \mid \mathcal{F}_n] - \mathbb{E}[e^{isZ}]\right|\right] \to 0.$$

This completes the proof by recalling (A.1) and (A.2). □

Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, USA
Email address: sayakc@wharton.upenn.edu

Department of Statistics, University of Chicago, USA
Email address: anirbanc@uchicago.edu

Department of Statistics, Columbia University, New York, USA
Email address: ac4662@columbia.edu

Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, USA
Email address: bhaswar@wharton.upenn.edu