



Contents lists available at ScienceDirect

European Journal of Combinatorics

journal homepage: www.elsevier.com/locate/ejc

Monochromatic subgraphs in randomly colored graphons

Bhaswar B. Bhattacharya^a, Sumit Mukherjee^{b,1}

^a Department of Statistics, University of Pennsylvania, Philadelphia, USA

^b Department of Statistics, Columbia University, NY, USA



ARTICLE INFO

Article history:

Received 29 June 2018

Accepted 10 June 2019

Available online xxxx

ABSTRACT

Let $T(H, G_n)$ be the number of monochromatic copies of a fixed connected graph H in a uniformly random coloring of the vertices of the graph G_n . In this paper we give a complete characterization of the limiting distribution of $T(H, G_n)$, when $\{G_n\}_{n \geq 1}$ is a converging sequence of dense graphs. When the number of colors grows to infinity, depending on whether the expected value remains bounded, $T(H, G_n)$ either converges to a finite linear combination of independent Poisson variables or a normal distribution. On the other hand, when the number of colors is fixed, $T(H, G_n)$ converges to a (possibly infinite) linear combination of independent centered chi-squared random variables. This generalizes the classical birthday problem, which involves understanding the asymptotics of $T(K_s, K_n)$, the number of monochromatic s -cliques in a complete graph K_n (s -matching birthdays among a group of n friends), to general monochromatic subgraphs in a network.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Let G_n be a simple labeled undirected graph with vertex set $V(G_n) := \{1, 2, \dots, |V(G_n)|\}$, edge set $E(G_n)$, and adjacency matrix $A(G_n) = \{a_{ij}(G_n), i, j \in V(G_n)\}$. In a *uniformly random c_n -coloring* of G_n , the vertices of G_n are colored with c_n colors as follows:

$$\mathbb{P}(v \in V(G_n) \text{ is colored with color } a \in \{1, 2, \dots, c_n\}) = \frac{1}{c_n}, \quad (1.1)$$

E-mail addresses: bhaswar@wharton.upenn.edu (B.B. Bhattacharya), sm3949@columbia.edu (S. Mukherjee).

¹ Research partially supported by National Science Foundation grant DMS-1712037.

<https://doi.org/10.1016/j.ejc.2019.06.007>

0195-6698/© 2019 Elsevier Ltd. All rights reserved.

independent from the other vertices. An edge $(a, b) \in E(G_n)$ is said to be *monochromatic* if $X_a = X_b$, where X_v denotes the color of the vertex $v \in V(G_n)$ in a uniformly random c_n -coloring of G_n . Denote by

$$T(K_2, G_n) = \frac{1}{2} \sum_{1 \leq u \neq v \leq |V(G_n)|} a_{uv}(G_n) \mathbf{1}\{X_u = X_v\}, \quad (1.2)$$

the number of monochromatic edges in G_n . Note that $\mathbb{P}(T(K_2, G_n) > 0) = 1 - \mathbb{P}(T(K_2, G_n) = 0) = 1 - \chi_{G_n}(c_n)/c_n^{|V(G_n)|}$, where $\chi_{G_n}(c_n)$ counts the number of proper colorings of G_n using c_n colors. The function χ_{G_n} is known as the *chromatic polynomial* of G_n , and is a central object in graph theory [18,22,23]. Moreover, the statistic (1.2) shows up in various applications, for example, in the study of coincidences [17] as a generalization of the birthday paradox [3,16], the Hamiltonian of the Ising/Potts models [4,8], and in non-parametric two-sample tests [5,20]. This requires understanding the asymptotics of $T(K_2, G_n)$ for various graph sequences G_n . The limiting distribution of $T(K_2, G_n)$ has been recently characterized by Bhattacharya et al. [7], for any sequence of growing graphs G_n .

Given these results, it is natural to ask what happens for monochromatic triangles, and more general subgraphs. In this paper we consider the problem of determining the limiting distribution of the number of monochromatic copies of a general connected simple graph H , in a uniformly random c_n -coloring of a graph sequence G_n . Formally, this is defined as

$$T(H, G_n) := \frac{1}{|Aut(H)|} \sum_{\mathbf{s} \in V(G_n)^{|V(H)|}} \prod_{(a,b) \in E(H)} a_{s_a s_b}(G_n) \mathbf{1}\{X_{\mathbf{s}}\},$$

where:

- $V(G_n)^{|V(H)|}$ is the set of all $|V(H)|$ -tuples $\mathbf{s} = (s_1, \dots, s_{|V(H)|}) \in V(G_n)^{|V(H)|}$ with distinct indices.² Thus, the cardinality of $V(G_n)^{|V(H)|}$ is $\frac{|V(G_n)|!}{(|V(G_n)| - |V(H)|)!}$.
- For any $\mathbf{s} = (s_1, \dots, s_{|V(H)|}) \in V(G_n)^{|V(H)|}$,

$$\mathbf{1}\{X_{\mathbf{s}}\} := \mathbf{1}\{X_{s_1} = \dots = X_{s_{|V(H)|}}\}. \quad (1.3)$$

- $Aut(H)$ is the *automorphism group* of H , that is, the number permutations σ of the vertex set $V(H)$ such that $(x, y) \in E(H)$ if and only if $(\sigma(x), \sigma(y)) \in E(H)$.

The class of possible limiting distributions of $T(H, G_n)$, for a general graph H , is extremely diverse [7], and, there appears to be no natural universality of the limiting distribution of $T(H, G_n)$, for a general graph sequence $\{G_n\}_{n \geq 1}$. In this paper, using results from the emerging theory of graph limits, we provide a complete characterization of the limiting distribution of $T(H, G_n)$, for any simple connected graph H , whenever $\{G_n\}_{n \geq 1}$ is a convergent sequence of dense graphs [24]. Depending on the behavior of $\mathbb{E}T(H, G_n)$ there are 3 different regimes:

- (1) $\mathbb{E}(T(H, G_n)) = O(1)$: In this case, $T(H, G_n)$ converges to a finite linear combination of independent Poisson random variables (Theorem 1.1).
- (2) $\mathbb{E}(T(H, G_n)) \rightarrow \infty$, such that $c_n \rightarrow \infty$: Here, $T(H, G_n)$ is asymptotically Gaussian, after appropriate standardization (Theorem 1.2).
- (3) $\mathbb{E}(T(H, G_n)) \rightarrow \infty$, such that $c_n = c$ is fixed: In this case, $T(H, G_n)$, after standardization, is asymptotically a (possibly infinite) linear combination of independent centered chi-squared random variables (Theorem 1.3).

We begin with a short background on graph limit theory. The results are formally stated in Section 1.2.

² For a set S , the set S^N denotes the N -fold cartesian product $S \times S \times \dots \times S$.

1.1. Graph limit theory

The theory of graph limits was developed by Lovász and coauthors [9,10,24], and has received phenomenal attention over the last few years. It builds a bridge between combinatorics and analysis, and has found applications in several disciplines including statistical physics, probability, and statistics [4,13,15]. For a detailed exposition of the theory of graph limits refer to Lovász [24]. Here we mention the basic definitions about the convergence of graph sequences. If F and G are two graphs, then define the homomorphism density of F into G by

$$t(F, G) := \frac{|\text{hom}(F, G)|}{|V(G)|^{|V(F)|}},$$

where $|\text{hom}(F, G)|$ denotes the number of homomorphisms of F into G . In fact, $t(F, G)$ is the proportion of maps $\phi : V(F) \rightarrow V(G)$ which define a graph homomorphism. Denote by $\text{hom}_{\text{inj}}(F, G)$ the number of injective maps from F into G which are homomorphisms, and

$$t_{\text{inj}}(F, G) := \frac{|\text{hom}_{\text{inj}}(F, G)|}{|V(G)|(|V(G)| - 1) \cdots (|V(G)| - |V(F)| + 1)},$$

which is the proportion of injective maps which are homomorphisms. Moreover, denote by $t_{\text{ind}}(F, G)$ the *induced homomorphism density*, that is, the proportion of injective maps $\phi : |V(F)| \rightarrow |V(G)|$, which satisfy $(\phi(x), \phi(y)) \in E(G)$ if and only if $(x, y) \in E(F)$:

$$t_{\text{ind}}(F, G) = \frac{\sum_{s \in (V(G) \setminus V(F))} \prod_{(a,b) \in E(F)} a_{sa} s_b(G) \prod_{(a,b) \notin E(F)} (1 - a_{sa} s_b(G))}{|V(G)|(|V(G)| - 1) \cdots (|V(G)| - |V(F)| + 1)}, \quad (1.4)$$

where $A(G) = (a_{ij}(G))_{i,j \in [|V(G)|]}$ is the adjacency matrix of G .

To define the continuous analogue of graphs, consider \mathcal{W} to be the space of all measurable functions from $[0, 1]^2$ into $[0, 1]$ that satisfy $W(x, y) = W(y, x)$, for all x, y . For a simple graph F with $V(F) = \{1, 2, \dots, |V(F)|\}$, let

$$t(F, W) = \int_{[0,1]^{|V(F)|}} \prod_{(i,j) \in E(F)} W(x_i, x_j) dx_1 dx_2 \cdots dx_{|V(F)|}.$$

Definition 1.1 ([9,10,24]). A sequence of graphs $\{G_n\}_{n \geq 1}$ is said to *converge to* W if for every finite simple graph F ,

$$\lim_{n \rightarrow \infty} t(F, G_n) = t(F, W). \quad (1.5)$$

If G_n converges to W , according to definition above, then the injective homomorphism densities converge: $t_{\text{inj}}(F, G_n) \rightarrow t(F, W)$, for every simple graph F . Moreover, the induced homomorphism densities also converge, that is, $t_{\text{ind}}(F, G_n) \rightarrow t_{\text{ind}}(F, W)$, for every simple graph F , where

$$t_{\text{ind}}(F, W) = \int_{[0,1]^{|V(F)|}} \prod_{(a,b) \in E(F)} W(x_a, x_b) \prod_{(a,b) \notin E(F)} (1 - W(x_a, x_b)) dx_1 dx_2 \cdots dx_{|V(F)|}. \quad (1.6)$$

The limit objects, that is, the elements of \mathcal{W} , are called *graph limits* or *graphons*. A finite simple graph $G = (V(G), E(G))$ can also be represented as a graphon in a natural way: Define $f^G(x, y) = \mathbf{1}\{(\lceil |V(G)|x \rceil, \lceil |V(G)|y \rceil) \in E(G)\}$, that is, partition $[0, 1]^2$ into $|V(G)|^2$ squares of side length $1/|V(G)|$, and let $f^G(x, y) = 1$ in the (i, j) -th square if $(i, j) \in E(G)$, and 0 otherwise. Observe that $t(F, f^G) = t(F, G)$ for every simple graph H and therefore the constant sequence G converges to the graph limit f^G . Define an equivalence relation on the space of graphons by $W_1 \sim W_2$ iff $t(F, W_1) = t(F, W_2)$ for all simple graphs F . It turns out that the quotient space under this equivalence relation, equipped with the notion of convergence in terms of subgraph densities outlined above is a compact metric space using the *cut distance* (refer to [24, Chapter 8]).

1.2. Results

Throughout the paper, we will assume that H is a finite, simple, and connected graph, and G_n is a sequence of dense graphs converging to the graphon W such that $t(H, W) > 0$. Depending on the limiting behavior of $\mathbb{E}T(H, G_n)$ there are 3 different regimes.

1.2.1. Linear combination of Poissons

For a finite simple unlabeled graph F , let $N(F, G_n)$ be the number of copies of F in G_n . Note that

$$N(H, G_n) = \frac{\text{hom}_{\text{inj}}(H, G_n)}{|Aut(H)|} \quad \text{and} \quad \mathbb{E}(T(H, G_n)) = \frac{N(H, G_n)}{c_n^{|V(H)|-1}}. \quad (1.7)$$

We begin with the regime where the mean $\mathbb{E}(T(H, G_n)) = O(1)$. In this case, the limit is a linear combination of independent Poisson variables, where the weights are determined by the limiting homomorphism densities of certain super-graphs of H . This is formalized in the following theorem:

Theorem 1.1. *Let G_n be a sequence of graphs converging to the graphon W , such that $t(H, W) > 0$. Suppose $c_n \rightarrow \infty$, such that $\mathbb{E}T(H, G_n) \rightarrow \lambda$. Then*

$$T(H, G_n) \xrightarrow{D} \sum_{F \supseteq H: |V(F)|=|V(H)|} N(H, F) X_F, \quad (1.8)$$

where $X_F \sim \text{Pois}\left(\lambda \cdot \frac{|Aut(H)|}{|Aut(F)|} \cdot \frac{t_{\text{ind}}(F, W)}{t(H, W)}\right)$ and the collection $\{X_F : F \supseteq H \text{ and } |V(F)| = |V(H)|\}$ is independent.³

The proof is based on a moment comparison technique, where the moments of $T(H, G_n)$ are compared with the moments of the corresponding random variable obtained by replacing every subset of $|V(H)|$ vertices with independent Bernoulli variables (refer to Section 2 for details).

Remark 1.1. A useful special case of the above theorem, which generalizes the well-known birthday problem, is when $H = K_s$ is the s -clique (monochromatic cliques correspond to s -matching birthdays in a friendship network G_n). The asymptotics of multiple birthday matches have found many applications, for example, in the study of coincidences [17, Problem 3], hash-function attacks in cryptology [25], and the discrete logarithm problem [6,21]. Refer to Example 2 for more on the birthday paradox.

Arratia et al. [2] used Stein's method based on dependency graphs to prove Poisson approximation theorems for $T(K_s, K_n)$, that is, the number of monochromatic s -cliques in a uniform coloring of a complete graph K_n (see also Chatterjee et al. [14]). Poisson limit theorems for $T(H, G_n)$ for general H and arbitrary coloring distribution are given in Cerquetti and Fortini [11]. They assumed that the distribution of colors was exchangeable and proved that $T(H, G_n)$ converges in distribution to a mixture of Poisson. However, these results only give conditions under which the limit of $T(H, G_n)$ is a Poisson, and require several assumptions on the number of certain subgraphs in G_n and the coloring distribution. On the other hand, Theorem 1.1 goes beyond the Poisson regime, and characterizes the limiting distribution of $T(H, G_n)$ for all dense graphs, under the uniform coloring distribution.

1.2.2. Asymptotic normality for growing colors

Theorem 1.1 asserts that if $\mathbb{E}T(H, G_n) \rightarrow \lambda$, then the number of monochromatic copies of H converges to a linear combination of Poissons. Recall that a Poisson random variable with mean growing to infinity converges to a standard normal distribution after centering by the mean and

³ For two graphs F and H , $F \supseteq H$ means F is a super-graph of H .

scaling by the standard deviation. Therefore, it is natural to wonder whether the same is true for $T(H, G_n)$, whenever $\mathbb{E}T(H, G_n) \rightarrow \infty$. To this end, define

$$Z(H, G_n) = \frac{T(H, G_n) - \mathbb{E}T(H, G_n)}{\sqrt{\text{Var}(T(H, G_n))}}. \quad (1.9)$$

The theorem shows that $Z(H, G_n)$ has a universal CLT whenever $\mathbb{E}T(H, G_n) \rightarrow \infty$ and $c_n \rightarrow \infty$. To this end, we define the Wasserstein distance between two probability measures μ and ν on \mathbb{R} is,

$$\text{Wass}(\mu, \nu) := \sup \left\{ \int f d\nu - \int f d\mu : f : \mathbb{R} \rightarrow \mathbb{R} \text{ is 1-Lipschitz} \right\},$$

that is, supremum over all f such that $|f(x) - f(y)| \leq |x - y|$. Moreover, for two nonnegative sequences $\{a_n\}_{n \geq 1}$ and $\{b_n\}_{n \geq 1}$, $a_n \lesssim b_n$ means $a_n \leq Cb_n$, for some constant $C > 0$ and all n large enough.

Theorem 1.2. Let G_n be a sequence of graphs converging to the graphon W , such that $t(H, W) > 0$. If $c_n \rightarrow \infty$, then

$$\text{Wass}(Z(H, G_n), N(0, 1)) \lesssim \left(\frac{c_n^{|V(H)|-1}}{|V(G_n)|^{|V(H)|}} \right)^{\frac{1}{2}} + \left(\frac{1}{c_n} \right)^{\frac{1}{2}}. \quad (1.10)$$

This implies, if $t(H, W) > 0$, then $Z(H, G_n) \xrightarrow{D} N(0, 1)$, whenever $\mathbb{E}T(H, G_n) \rightarrow \infty$ and $c_n \rightarrow \infty$.

The proof of the above theorem is given in Section 3, and is based on a Stein's method based on dependency graphs.

Remark 1.2. For the case of monochromatic edges, [7, Theorem 1.2] showed that $Z(K_2, G_n) \xrightarrow{D} N(0, 1)$, whenever $\mathbb{E}(T(K_2, G_n)) \rightarrow \infty$ such that $c_n \rightarrow \infty$, for any sequence of graphs G_n with $|V(G_n)| \rightarrow \infty$. Error rates for the above CLT were obtained by Fang [19]. The above theorem shows that this phenomenon extends to all simple connected graphs H , when G_n is a converging sequence of dense graphs. Moreover, unlike in the case of edges, the density assumption $t(H, W) > 0$ is, in general, necessary for $Z(H, G_n)$ to have a non-degenerate normal limit (see Example 4).

1.2.3. Limiting distribution for fixed number of colors

In this section we derive the asymptotic distribution of the number of monochromatic subgraphs when $\mathbb{E}T(H, G_n) \rightarrow \infty$ such that c is fixed.

Definition 1.2 (2-Point Homomorphism Functions for Graphons). Let H be a labeled finite simple graph and W is a graphon. Then, for $1 \leq u \neq v \leq |V(H)|$, the function $t_{u,v}(\cdot, \cdot, H, W) : [0, 1]^2 \rightarrow [0, 1]$ is defined as:⁴

$$t_{u,v}(x, y, H, W) = W_{u,v}^+(x, y) \int_{[0,1]^{|V(H)|-2}} \prod_{r \in N_H(u) \setminus \{v\}} W(x, z_r) \prod_{s \in N_H(v) \setminus \{u\}} W(y, z_s) \prod_{(r,s) \in E(H) \setminus \{u,v\}} W(z_r, z_s) \prod_{r \notin \{u,v\}} dz_r,$$

with $W_{u,v}^+(x, y) = W(x, y)$ if $(u, v) \in E(H)$ and 1 otherwise. Note that $t_{v,u}(x, y, H, W) = t_{u,v}(y, x, H, W)$.

For example, when $H = K_{1,2}$ is the 2-star, with the central vertex labeled 1. Then the following hold:

⁴ For a graph $F = (V(F), E(F))$ and $S \subseteq V(F)$, the neighborhood of S in F is $N_F(S) = \{v \in V(F) : \exists u \in S \text{ such that } (u, v) \in E(F)\}$. Moreover, for $u, v \in V(F)$, $F \setminus \{u, v\}$ is the graph obtained by removing the vertices u, v and all the edges incident on them.

- $t_{1,2}(x, y, K_{1,2}, W) = t_{1,3}(x, y, K_{1,2}, W) = W(x, y)d_W(x)$, where $d_W(x) = \int_0^1 W(x, z)dz$ is the degree function of the graphon W .
- $t_{2,3}(x, y, K_{1,2}, W) = \int_{[0,1]} W(x, z_1)W(y, z_1)dz_1$.

Similarly, $t_{2,1}(x, y, K_{1,2}, W) = t_{3,1}(x, y, K_{1,2}, W) = W(x, y)d_W(y)$, and

$$t_{3,2}(x, y, K_{1,2}, W) = \int_{[0,1]} W(x, z_1)W(y, z_1)dz_1.$$

More examples are computed in Section 4.2.

Using this definition we can now show that the limiting distribution of

$$\Gamma(H, G_n) = \frac{T(H, G_n) - \mathbb{E}T(H, G_n)}{|V(G_n)|^{|V(H)|-1}}, \quad (1.11)$$

is a linear combination of centered chi-squared random variables, whenever $\{G_n\}_{n \geq 1}$ converges and the number of colors is fixed. To this end, note that every bounded non-negative symmetric function $K : [0, 1]^2 \rightarrow \mathbb{R}$ defines an operator $T_K : L_2[0, 1] \rightarrow L_2(\mathbb{R})$, by

$$(T_K f)(x) = \int_0^1 K(x, y)f(y)dy.$$

T_K is a Hilbert–Schmidt operator, which is compact and has a discrete spectrum, that is, a countable multi-set of non-zero real eigenvalues $\{\lambda_1(K), \lambda_2(K), \dots\}$, where every non-zero eigenvalue has finite multiplicity.

Theorem 1.3. *Let G_n be a sequence of graphs converging to the graphon W , such that $t(H, W) > 0$. If $c_n = c$ is fixed, then*

$$\Gamma(H, G_n) \xrightarrow{D} \frac{1}{c^{|V(H)|-1}} \cdot \sum_{r=1}^{\infty} \lambda_r(H, W) \cdot \eta_r, \quad (1.12)$$

where

- (η_1, η_2, \dots) is a collection of independent $\chi_{(c-1)}^2 - (c-1)$ random variables,⁵
- $\{\lambda_1(H, W), \lambda_2(H, W), \dots\}$ is the multi-set of the non-zero eigenvalues of the bounded non-negative symmetric function $W_H : [0, 1]^2 \rightarrow \mathbb{R}$ defined by

$$W_H(x, y) := \frac{1}{2|Aut(H)|} \sum_{1 \leq u \neq v \leq |V(H)|} t_{u,v}(x, y, H, W). \quad (1.13)$$

(If the spectrum of W_H is finite, then the sum in (1.12) is interpreted by padding zeros to the spectrum.)

Note that the function W_H is symmetric because of the relation $t_{v,u}(x, y, H, W) = t_{u,v}(y, x, H, W)$ and is point-wise bounded by $W_H \leq \frac{|V(H)|^2}{2|Aut(H)|}$. The proof of the above theorem is given in Section 4. It has two main steps:

- The first step expands the random variable $\Gamma(H, G_n)$ as a polynomial in the i.i.d. color vectors $\{\mathbf{1}\{X_v = a\}\}_{a \in [c]} : v \in V(G_n)\}$, and shows that only the quadratic term is dominant (refer to Lemma 4.1 and Lemma 4.2 for details).
- The second step shows that the limiting distribution of the quadratic term remains unchanged when the color vectors are replaced by a collection of i.i.d. Gaussian random vectors with the same mean and covariance structure (Lemma 4.3). The result then follows by analyzing the asymptotics of the Gaussian counterpart.

⁵ Recall that for a positive integer a , χ_a^2 denotes the chi-squared distribution with a degrees of freedom, that is, the sum of squares of a independent standard normal $(N(0, 1))$ random variables.

1.3. Organization

The rest of the paper is organized as follows: The proof of [Theorem 1.1](#) and its applications are given in Section 2. The proof of [Theorem 1.2](#) is in Section 3. The proof of [Theorem 1.3](#) and related examples are discussed in Section 4.

2. Linear combination of Poissons: Proof of [Theorem 1.1](#)

In this section we present the proof of [Theorem 1.1](#) and discuss applications of this result in various examples.

2.1. Proof of [Theorem 1.1](#)

To analyze $T(H, G_n)$ we use the ‘independent approximation’, where the indicators $\mathbf{1}\{X_{\mathbf{s}}\}$ are replaced by independent Bernoulli variables, for every subset of vertices in G_n of size $|V(H)|$. To this end, define

$$J(H, G_n) = \frac{1}{|Aut(H)|} \sum_{\mathbf{s} \in V(G_n)^{|V(H)|}} \prod_{(a,b) \in E(H)} a_{s_a s_b} J_{\mathbf{s}}, \quad (2.1)$$

where $\{J_{\mathbf{s}}, s_1 < s_2 < \dots < s_{|V(H)|} \in V(G_n)\}$ is a collection of i.i.d. $\text{Ber}(\frac{1}{c_n^{|V(H)|-1}})$ random variables, and if the coordinates of \mathbf{s} are not in increasing order, define $J_{\mathbf{s}} = J_{\sigma(\mathbf{s})}$, where $\sigma(\mathbf{s}) = (\sigma(s_1), \sigma(s_2), \dots, \sigma(s_{|V(H)|}))$ such that $\sigma(s_1) < \sigma(s_2) < \dots < \sigma(s_{|V(H)|})$. This implies that

$$\{J_{\mathbf{s}}, \mathbf{s} \in \binom{V(G_n)}{|V(H)|}\} \quad (2.2)$$

is a collection of i.i.d. $\text{Ber}(\frac{1}{c_n^{|V(H)|-1}})$ random variables, where $\binom{V(G_n)}{|V(H)|}$ denotes the set of all $|V(H)|$ -element subset of $V(G_n)$.

The following lemma shows that the moments of $T(H, G_n)$ and $J(H, G_n)$ are asymptotically close. Note that $A \lesssim_H B$, means $A \leq C(H)B$, where $C(H)$ is a constant that depends only on the graph H . Similarly, $A \gtrsim_H B$, means $A \geq C(H)B$, where $C(H)$ is a constant that depends only on the graph H .

Lemma 2.1. For any $r \geq 1$,

$$\lim_{n \rightarrow \infty} \left\{ \mathbb{E}T(H, G_n)^r - \mathbb{E}J(H, G_n)^r \right\} = 0. \quad (2.3)$$

Moreover, there exists a constant $C = C(H, r) < \infty$ such that for all n large, $\mathbb{E}T(H, G_n)^r \leq C$ and $\mathbb{E}J(H, G_n)^r \leq C$.

Proof. Let \mathcal{A} be the collection of all ordered r -tuples $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_r$, where $\mathbf{s}_j \in V(G_n)^{|V(H)|}$ for $j \in [r]$, with $\mathbf{s}_1 = (s_{11}, s_{12}, \dots, s_{1|V(H)|})$, $\mathbf{s}_2 = (s_{21}, s_{22}, \dots, s_{2|V(H)|})$, \dots , $\mathbf{s}_r = (s_{r1}, s_{r2}, \dots, s_{r|V(H)|})$, such that $\prod_{(a,b) \in E(H)} a_{s_{ja} s_{jb}} = 1$, for every $j \in [r]$. Then by the multinomial expansion,

$$\begin{aligned} |\mathbb{E}T(H, G_n)^r - \mathbb{E}J(H, G_n)^r| &\leq \frac{1}{|Aut(H)|^r} \sum_{\mathcal{A}} \left| \mathbb{E} \prod_{t=1}^r \mathbf{1}\{X_{\mathbf{s}_t}\} - \mathbb{E} \prod_{t=1}^r J_{\mathbf{s}_t} \right| \\ &= \frac{1}{|Aut(H)|^r} \sum_{\mathcal{A}} \left| \frac{1}{c_n^{|V(F)|-v(F)}} - \frac{1}{c_n^{b|V(H)|-b}} \right|, \end{aligned}$$

where $F = F(\mathbf{s}_1, \dots, \mathbf{s}_r)$ is the graph on vertex set $V(F) = \bigcup_{t=1}^r \mathbf{s}_t$ and edge set $\bigcup_{t=1}^r \{(s_{ta}, s_{tb}) : (a, b) \in E(H)\}$, $v(F)$ is the number of connected components of F , and b is the number of distinct

$|V(H)|$ -element subsets in the collection $\{s_1, s_2, \dots, s_r\}$.⁶ Note that if the graph F is connected, $|V(F)| - 1 \leq b|V(H)| - b$, and therefore, in general $|V(F)| - v(F) \leq b|V(H)| - b$.

We now claim that $|V(F)| - v(F) < b|V(H)| - b$ implies $|V(F)| > |V(H)|v(F)$. Indeed, first note that trivially $|V(F)| \geq |V(H)|v(F)$. If $|V(F)| = |V(H)|v(F)$, then every connected component of F is isomorphic H , that is, $v(F) = b$ and $|V(F)| - v(F) = b|V(H)| - b$, verifying the claim. Thus, setting $\mathcal{N}_{p,q,r}$ to be the set of all r ordered tuples s_1, \dots, s_r in \mathcal{A} such that $|\bigcup_{t=1}^r s_t| = p$ and $v(F) = q$, we have

$$\begin{aligned} |\mathbb{E}T(H, G_n)^r - \mathbb{E}J(H, G_n)^r| &\lesssim_H \sum_{(p,q): q|V(H)| < p \leq r|V(H)|} \sum_{\mathcal{N}_{p,q,r}} \frac{1}{c_n^{p-q}} \\ &\lesssim_H \sum_{(p,q): q|V(H)| < p \leq r|V(H)|} \frac{|V(G_n)|^p}{c_n^{p-q}} \quad (\text{using } |\mathcal{N}_{p,q,r}| = O(|V(G_n)|^p)) \\ &\lesssim_H \sum_{(p,q): q|V(H)| < p \leq r|V(H)|} \frac{|V(G_n)|^p}{|V(G_n)|^{\frac{|V(H)|}{|V(H)|-1}(p-q)}} \\ &\quad (\text{since } c_n^{|V(H)|-1} = \Theta(|V(G_n)|^{|V(H)|})) \\ &\lesssim_H \sum_{(p,q): q|V(H)| < p \leq r|V(H)|} \frac{1}{|V(G_n)|^{\frac{1}{|V(H)|-1}(p-q|V(H)|)}}. \end{aligned}$$

Since $p > q|V(H)|$, each term in the above sum converges to 0, and because the sum is over a finite index set free of n , (2.3) follows.

Finally, from the above arguments it also follows that

$$\mathbb{E}T(H, G_n)^r \lesssim_H \sum_{(p,q): q|V(H)| \leq p \leq r|V(H)|} \frac{1}{|V(G_n)|^{\frac{1}{|V(H)|-1}(p-|V(H)|q)}} = O(1),$$

since $p \geq |V(H)|q$, for all (p, q) in the above sum. \square

Next, we show that the limiting distribution of $J(H, G_n)$ is a linear combination of independent Poisson random variables.

Lemma 2.2. Let $J(H, G_n)$ be as defined in (2.1). Then

$$J(H, G_n) \rightarrow \sum_{F \supseteq H: |V(F)|=|V(H)|} N(H, F)X_F,$$

in distribution and in moments, where $X_F \sim \text{Pois}\left(\lambda \cdot \frac{|Aut(H)|}{|Aut(F)|} \cdot \frac{t_{ind}(F, W)}{t(H, W)}\right)$ and the collection $\{X_F : F \supseteq H \text{ and } |V(F)| = |V(H)|\}$ is independent.

Proof. Let $\binom{V(G_n)}{|V(H)|}$ be the collection of $|V(H)|$ -element subsets of $V(G_n)$. For $S \subseteq V(G_n)$ denote by $G_n[S]$ the subgraph of G_n induced on the set S . Then recalling the definition of $J(H, G_n)$ from (2.1) gives

$$\begin{aligned} J(H, G_n) &= \frac{1}{|Aut(H)|} \sum_{s \in \binom{V(G_n)}{|V(H)|}} \prod_{(a,b) \in E(H)} a_{s_a s_b}(G_n) j_s \\ &= \sum_{s \in \binom{V(G_n)}{|V(H)|}} N(H, G_n[s]) j_s \end{aligned}$$

⁶ Here, an ordered $|V(H)|$ -tuple $\mathbf{z} = (z_1, z_2, \dots, z_{|V(H)|}) \in V(G_n)_{|V(H)|}$ is considered as a $|V(H)|$ -element subset $\{z_1, z_2, \dots, z_{|V(H)|}\}$ of $V(G_n)$. For example, in the collection of ordered triples $\{(1, 2, 3), (3, 2, 1), (4, 3, 6), (5, 4, 1), (3, 6, 4), (5, 6, 8)\}$, there are 4 distinct 3-element subsets, that is, $b = 4$. Hereafter, we will slightly abuse notation and use the tuple and set interpretations interchangeably, whenever it is clear from the context.

$$= \sum_{F \supseteq H: |V(F)|=|V(H)|} N(H, F) \sum_{\mathbf{s} \in \binom{V(G_n)}{|V(H)|}} \mathbf{1}\{G_n[\mathbf{s}] = F\} \cdot J_{\mathbf{s}}. \quad (2.4)$$

Now, note that, by (2.2), the collection

$$\left\{ \sum_{\mathbf{s} \in \binom{V(G_n)}{|V(H)|}} \mathbf{1}\{G_n[\mathbf{s}] = F\} \cdot J_{\mathbf{s}} : F \supseteq H \text{ and } |V(F)| = |V(H)| \right\}$$

is independent, since for any two distinct super-graphs $F_1, F_2 \supseteq H$, with $|V(F_1)| = |V(F_2)| = |V(H)|$, the sets

$$\left\{ \mathbf{s} \in \binom{V(G_n)}{|V(H)|} : \mathbf{1}\{G_n[\mathbf{s}] = F_1\} \right\}, \quad \text{and} \quad \left\{ \mathbf{s} \in \binom{V(G_n)}{|V(H)|} : \mathbf{1}\{G_n[\mathbf{s}] = F_2\} \right\}$$

are disjoint. Moreover, for every fixed F ,

$$J_n(F) := \sum_{\mathbf{s} \in \binom{V(G_n)}{|V(H)|}} \mathbf{1}\{G_n[\mathbf{s}] = F\} J_{\mathbf{s}}$$

is a sum of independent $\text{Ber}\left(\frac{1}{c_n^{|V(H)|-1}}\right)$ random variables. Therefore, to prove theorem it suffices to show that $J_n(F) \rightarrow X_F$ (with X_F as defined in the statement of the theorem) in distribution and in moments, which follows if we can prove that

$$\mathbb{E}(J_n(F)) \rightarrow \lambda \cdot \frac{|Aut(H)|}{|Aut(F)|} \cdot \frac{t_{\text{ind}}(F, W)}{t(H, W)}. \quad (2.5)$$

To show (2.5), first note that

$$\frac{|V(G_n)|^{|V(H)|}}{c_n^{|V(H)|-1}} = \frac{N(H, G_n)}{c_n^{|V(H)|-1}} \cdot \frac{|V(G_n)|^{|V(H)|}}{N(H, G_n)} = (1 + o(1))\lambda \cdot \frac{|V(G_n)|^{|V(H)|}}{\text{hom}_{\text{inj}}(H, G_n)/|Aut(H)|} \rightarrow \lambda \frac{|Aut(H)|}{t(H, W)},$$

using $\frac{\text{hom}_{\text{inj}}(H, G_n)}{|V(G_n)|^{|V(H)|}} = (1 + o(1))t_{\text{inj}}(H, G_n) \rightarrow t(H, W)$ (by [24, (5.21)]).

Then recalling (1.7),

$$\begin{aligned} & \mathbb{E} \sum_{\mathbf{s} \in \binom{V(G_n)}{|V(H)|}} \mathbf{1}\{G_n[\mathbf{s}] = F\} \cdot J_{\mathbf{s}} \\ &= \frac{1}{c_n^{|V(H)|-1}} \sum_{\mathbf{s} \in \binom{V(G_n)}{|V(H)|}} \prod_{(a,b) \in E(F)} a_{s_a s_b}(G_n) \prod_{(a,b) \notin E(F)} (1 - a_{s_a s_b}(G_n)) \\ &= \frac{|V(G_n)|^{|V(H)|}}{c_n^{|V(H)|-1}} \frac{1}{|V(G_n)|^{|V(H)|}} \sum_{\mathbf{s} \in \binom{V(G_n)}{|V(F)|}} \prod_{(a,b) \in E(F)} a_{s_a s_b}(G_n) \prod_{(a,b) \notin E(F)} (1 - a_{s_a s_b}(G_n)) \\ &= (1 + o(1))\lambda \frac{|Aut(H)|}{t(H, W)} \cdot \frac{1}{|V(G_n)|^{|V(H)|}} \sum_{\mathbf{s} \in \binom{V(G_n)}{|V(F)|}} \prod_{(a,b) \in E(F)} a_{s_a s_b}(G_n) \prod_{(a,b) \notin E(F)} (1 - a_{s_a s_b}(G_n)) \\ &= (1 + o(1))\lambda \frac{|Aut(H)|}{t(H, W)} \cdot \frac{t_{\text{ind}}(F, G_n)}{|Aut(F)|} \quad (\text{recall (1.4)}) \\ &\rightarrow \lambda \frac{|Aut(H)|}{t(H, W)} \cdot \frac{t_{\text{ind}}(F, W)}{|Aut(F)|}. \end{aligned} \quad (2.6)$$

where the last step uses $t_{\text{ind}}(F, G_n) \rightarrow t_{\text{ind}}(F, W)$ (since G_n converges to W). \square

The proof of Theorem 1.1 can be easily completed using the above two lemmas: Let $Y := \sum_{F \supseteq H: |V(F)|=|V(H)|} N(H, F) X_F$. By Lemmas 2.1 and 2.2, $T(H, G_n)$ converges in moments to Y . Now, it is

easy to check that Y satisfies the Stieltjes moment condition [1], therefore, it is uniquely determined by its moments. This implies $T(H, G_n) \xrightarrow{D} Y$ as well, and hence completes the proof of Theorem 1.1.

2.2. Examples

Theorem 1.1 can be easily extended to converging sequence of dense random graphs, when the limits in (1.5) hold in probability, by conditioning on the graph, under the assumption that the graph and its coloring are jointly independent. Here, we compute the limiting distribution (1.8) for the Erdős–Rényi random graph.

Example 1 (Erdős–Rényi Random Graphs). Let $G_n \sim G(n, p)$ be the Erdős–Rényi random graph. In this case G_n converges to the constant function $W^{(p)} = p$. This implies that $t(H, W^{(p)}) = p^{|E(H)|}$ and $t_{\text{ind}}(F, W^{(p)}) = p^{|E(F)|} (1-p)^{\binom{|V(H)|}{2} - |E(F)|}$. Therefore, by Theorem 1.1, choosing c_n such that $\mathbb{E}T(H, G_n) = (1 + o(1)) \frac{\binom{|V(G_n)|}{|V(H)|} p^{|E(H)|}}{c_n^{|V(H)|-1}} \rightarrow \lambda$, gives

$$T(H, G_n) \xrightarrow{D} \sum_{F \supseteq H: |V(F)|=|V(H)|} N(H, F) X_F,$$

where $X_F \sim \text{Pois}\left(\lambda \cdot \frac{|Aut(H)|}{|Aut(F)|} p^{|E(F)|-|E(H)|} (1-p)^{\binom{|V(H)|}{2}-|E(F)|}\right)$ and the collection $\{X_F : F \supseteq H \text{ and } |V(F)| = |V(H)|\}$ is independent.

- When $G_n = K_n$ (that is, $p = 1$), $X_{K_{|V(H)|}} \sim \text{Pois}\left(\lambda \cdot \frac{|Aut(H)|}{|V(H)|!}\right) = \text{Pois}\left(\frac{\lambda}{N(H, K_{|V(H)|})}\right)$ and $X_F = 0$ otherwise. Therefore,

$$T(H, K_n) \xrightarrow{D} N(H, K_{|V(H)|}) \cdot \text{Pois}\left(\frac{\lambda}{N(H, K_{|V(H)|})}\right).$$

- If $H = K_s$ is the complete graph, then $\{F \supseteq H : |V(F)| = |V(H)|\} = \{H\}$ and, therefore, $T(H, G_n) \xrightarrow{D} \text{Pois}(\lambda)$.

When $H = K_s$ is the s -clique, we have a birthday problem on a general friendship network G_n .

Example 2 (Birthday Problem). In the well-known birthday problem, G_n is a friendship-network graph where the vertices are colored uniformly with $c_n = 365$ colors (corresponding to birthdays). In this case, two friends will have the same birthday whenever the corresponding edge in the graph G_n is monochromatic. Therefore, $\mathbb{P}(T(K_s, G_n) > 0)$ is the probability that there is an s -fold birthday match, that is, there are s friends with the same birthday. For this problem, Theorem 1.1 can be used to do an approximate sample size calculation. For example, using $T(K_s, G_n) \xrightarrow{D} \text{Pois}(\lambda)$, where $\frac{N(K_s, G_n)}{c_n^{s-1}} \rightarrow \lambda$, and $\frac{1}{|V(G_n)|^s} N(K_s, G_n) \rightarrow \frac{1}{s!} t(K_s, W)$, gives

$$\mathbb{P}(T(K_s, G_n) > 0) \approx 1 - e^{-\frac{N(K_s, G_n)}{c_n^{s-1}}} = p,$$

$$\text{which implies } |V(G_n)| \approx \left(\frac{s!}{t(K_s, W)} c_n^{s-1} \log\left(\frac{1}{1-p}\right) \right)^{\frac{1}{s}},$$

which approximates the minimum number of people needed to ensure a s -fold birthday match in the network G_n , with probability at least p . When the underlying graph $G_n = K_n$ is the complete graph K_n on n vertices, this reduces to the classical birthday problem. For example, when $G_n = K_n$, $p = \frac{1}{2}$ and $s = 3$, using $c_n = 365$, the RHS above evaluates approximately to 82.1, that is, in any group of 83 people, with probability at least 50%, there are three friends all having the same birthday.

The assumption $t(H, W) > 0$ in Theorem 1.1 enforces that $c_n^{|V(H)|-1} = \Theta(|V(G_n)|^{|V(H)|})$, and, in this regime, the limiting distribution of $T(H, G_n)$ is a linear combination of Poissons (1.8). However

when $t(H, W) = 0$, for the scaling $c_n^{|V(H)|-1} \sim N(H, G_n)$ we can get ‘non-linear’ limiting distributions, as shown below.

Example 3 (Product of Independent Poissons). Let $G_n = K_{1,n,n}$, the complete 3-partite graph, with partite sets $\{z\}, B, C$ such that $|B| = |C| = n$. Note that every triangle in G_n passes through z , hence, $N(K_3, G_n) = n^2$. In this case, the limiting graphon is $W(x, y) = \mathbf{1}\{(x - \frac{1}{2})(y - \frac{1}{2}) \leq 0\}$, for which $t(K_3, W) = 0$. However, if we color G_n randomly with $c_n = n$ colors, such that $N(K_3, G_n)/c_n^2 = 1$, then $T(K_3, G_n)$ has a non-degenerate limiting distribution: For $a \in [c_n]$, let $L_n(a)$ and $R_n(a)$ be the number of vertices in sets B and C with color a , respectively. Clearly,

$$L_n(a) \sim \text{Bin}(n, 1/c_n) \xrightarrow{D} X, \quad R_n(a) \sim \text{Bin}(n, 1/c_n) \xrightarrow{D} Y,$$

where X and Y are independent $\text{Pois}(1)$ variables. Thus, given the color of the vertex z is a , $T(K_3, G_n) = L_n(a)R_n(a) \xrightarrow{D} XY$, and therefore, unconditionally $T(K_3, G_n) \xrightarrow{D} XY$, the product of two independent $\text{Pois}(1)$ random variables.

3. Asymptotic normality: Proof of Theorem 1.2

In this section we prove the asymptotic normality of $Z(H, G_n)$, whenever $t(H, W) > 0$ and $\mathbb{E}T(H, G_n) \rightarrow \infty$ such that $c_n \rightarrow \infty$. We begin the following definition.

Definition 3.1. Given a graph H with vertices labeled $\{1, 2, \dots, |V(H)|\}$ such that $a, b \in V(H)$, define the (a, b) -join of H , denoted by $H_{(a,b)}^2$ as follows: Let H' be an isomorphic copy of H with vertices $\{1', 2', \dots, |V(H)|'\}$, where the vertex s maps to the vertex s' , for $s \in V(H)$. The graph $H_{(a,b)}^2$ is obtained by identifying the vertex a and b in H with the vertex a' and b' in H' , that is, $V(H_{(a,b)}^2) = V(H) \cup V(H') \setminus \{a', b'\}$ and

$$E(H_{(a,b)}^2) = E(H) \cup E(H' \setminus \{a', b'\}) \cup \{(a, x') : x' \in N_{H'}(a')\} \cup \{(b, y') : y' \in N_{H'}(b')\}.$$

Note that $H_{(a,b)}^2$ has $2|V(H)| - 2$ vertices and $2|E(H)| - 1$ or $2|E(H)|$ edges, depending on whether the edge (a, b) is present or absent in $E(H)$ respectively.

Lemma 3.1. Let H be a graph with vertices labeled $\{1, 2, \dots, |V(H)|\}$ such that $a, b \in V(H)$. Then $t(H_{(a,b)}^2, W) > 0$, whenever $t(H, W) > 0$.

Proof. Recalling Definition 1.2, gives

$$\begin{aligned} t(H, W) &= \int_{[0,1]^2} t_{a,b}(x_a, x_b, H, W) dx_a dx_b \\ &\leq \left(\int_{[0,1]^2} t_{a,b}^2(x_a, x_b, H, W) dx_a dx_b \right)^{\frac{1}{2}} = (t(H_{(a,b)}^2, W))^{\frac{1}{2}} \quad (\text{by Cauchy-Schwarz}) \end{aligned}$$

This implies $t(H_{(a,b)}^2, W) \geq t(H, W)^2 > 0$. \square

Definition 3.2. For $\mathbf{s} \in V(G_n)_{|V(H)|}$, define

$$Z_{\mathbf{s}} := \mathbf{1}\{X_{=\mathbf{s}}\} - \mathbb{E}\mathbf{1}\{X_{=\mathbf{s}}\} = \mathbf{1}\{X_{=\mathbf{s}}\} - \frac{1}{c_n^{|V(H)|-1}}, \quad (3.1)$$

where $\mathbf{1}\{X_{=\mathbf{s}}\}$ is as defined in (1.3). Then

$$\begin{aligned} T(H, G_n) - \mathbb{E}T(H, G_n) &= \sum_{\mathbf{s} \in V(G_n)_{|V(H)|}} M_{G_n}(\mathbf{s}, H) \left(\mathbf{1}\{X_{=\mathbf{s}}\} - \frac{1}{c_n^{|V(H)|-1}} \right) \\ &= \sum_{\mathbf{s} \in V(G_n)_{|V(H)|}} M_{G_n}(\mathbf{s}, H) Z_{\mathbf{s}}, \end{aligned} \quad (3.2)$$

where $M_{G_n}(\mathbf{s}, H) = \frac{1}{|Aut(H)|} \prod_{(a,b) \in E(H)} a_{s_a s_b}(G_n)$.

The following lemma calculates the covariance of Z_s and Z_t and obtains a lower bound on the variance of $T(H, G_n)$.

Lemma 3.2. *The following hold:*

(a) For $\mathbf{s}, \mathbf{t} \in V(G_n)_{|V(H)|}$,

$$\mathbb{E}\mathbf{1}\{X_{=\mathbf{s}}\}\mathbf{1}\{X_{=\mathbf{t}}\} = \begin{cases} \frac{1}{c_n^{|s \cup t|-2}} & \text{if } |s \cup t| = 2|V(H)| \\ \frac{1}{c_n^{|s \cup t|-1}} & \text{if } |s \cup t| \leq 2|V(H)| - 1. \end{cases}$$

(b) For $\mathbf{s}, \mathbf{t} \in V(G_n)_{|V(H)|}$,

$$\mathbb{E}Z_s Z_t = \begin{cases} \frac{1}{c_n^{|s \cup t|-1}} - \frac{1}{c_n^{2|V(H)|-2}} & \text{if } |s \cup t| \leq 2|V(H)| - 2 \\ 0 & \text{if } |s \cup t| \in \{2|V(H)| - 1, 2|V(H)|\}. \end{cases}$$

(c) If $t(H, W) > 0$, then

$$\text{Var}(T(H, G_n)) \gtrsim \max \left(\frac{|V(G_n)|^{|V(H)|}}{c_n^{|V(H)|-1}}, \frac{|V(G_n)|^{2|V(H)|-2}}{c_n^{2|V(H)|-3}} \right).$$

Proof. If $|s \cup t| = 2|V(H)|$, the indices \mathbf{s}, \mathbf{t} do not intersect. In this case, the expectation factorizes by independence, and

$$\mathbb{E}\mathbf{1}\{X_{=\mathbf{s}}\}\mathbf{1}\{X_{=\mathbf{t}}\} = \mathbb{E}\mathbf{1}\{X_{=\mathbf{s}}\}\mathbb{E}\mathbf{1}\{X_{=\mathbf{t}}\} = \frac{1}{c_n^{2|V(H)|-2}}.$$

Otherwise,

$$\mathbb{E}\mathbf{1}\{X_{=\mathbf{s}}\}\mathbf{1}\{X_{=\mathbf{t}}\} = \mathbb{P}(X_{s_1} = \dots = X_{s_{|V(H)|}} = X_{t_1} = \dots = X_{t_{|V(H)|}}) = \frac{1}{c_n^{|s \cup t|-1}},$$

completing the proof of (a). The result in (b) follows from (a) and observing that $\mathbb{E}\mathbf{1}\{X_{=\mathbf{s}}\}\mathbb{E}\mathbf{1}\{X_{=\mathbf{t}}\} = \frac{1}{c_n^{2|V(H)|-2}}$.

To show (c) note that, by (3.2),

$$\begin{aligned} \text{Var}(T(H, G_n)) &= \sum_{\mathbf{s} \in V(G_n)_{|V(H)|}} \frac{M_{G_n}(\mathbf{s}, H)}{|Aut(H)|} \mathbb{E}Z_s^2 + \sum_{\mathbf{s} \neq \mathbf{t} \in V(G_n)_{|V(H)|}} M_{G_n}(\mathbf{s}, H) M_{G_n}(\mathbf{t}, H) \mathbb{E}Z_s Z_t, \end{aligned} \quad (3.3)$$

using $M_{G_n}(\mathbf{s}, H)^2 = \frac{1}{|Aut(H)|^2} \prod_{(a,b) \in E(H)} a_{s_a s_b}(G_n) = \frac{1}{|Aut(H)|} M_{G_n}(\mathbf{s}, H)$. Now, since each of terms in the covariance is non-negative by part (b),

$$\text{Var}(T(H, G_n)) \gtrsim_H \sum_{\mathbf{s} \in V(G_n)_{|V(H)|}} M_{G_n}(\mathbf{s}, H) \mathbb{E}Z_s^2 \gtrsim_H \frac{N(H, G_n)}{c_n^{|V(H)|-1}} \gtrsim \frac{|V(G_n)|^{|V(H)|}}{c_n^{|V(H)|-1}},$$

since $\frac{N(H, G_n)}{|V(G_n)|^{|V(H)|}} = (1 + o(1))^{\frac{t_{\text{inj}}(H, G_n)}{|Aut(H)|}} \rightarrow \frac{t(H, W)}{|Aut(H)|} > 0$, when G_n converges to W .

Another way to lower bound $\text{Var}(T(H, G_n))$ is to use that the first term in (3.3) is non-negative, and to consider only the sum over indices $|s \cap t| = 2$ in the second term. The sum over such pairs $\mathbf{s}, \mathbf{t} \in V(G_n)_{|V(H)|}$, that is, $\mathbf{s} = (s_1, s_2, s_3, \dots, s_{|V(H)|})$, $\mathbf{t} = (s_1, s_2, t_3, \dots, t_{|V(H)|})$ with the indices $\{s_1, s_2, s_3, \dots, s_{|V(H)|}, t_3, \dots, t_{|V(H)|}\}$ all distinct gives

$$\text{Var}(T(H, G_n)) \gtrsim_H \frac{N(H_{(1,2)}^2, G_n)}{c_n^{2|V(H)|-3}} \gtrsim \frac{|V(G_n)|^{2|V(H)|-2}}{c_n^{2|V(H)|-3}},$$

where the last step uses $\lim_{n \rightarrow \infty} \frac{1}{|V(G_n)|^{2|V(H)|-2}} N(H_{(1,2)}^2, G_n) \gtrsim t(H_{(1,2)}^2, W) > 0$ by Lemma 3.1. Combining these two estimates gives the desired lower bound on the variance. \square

In the following two lemmas we estimate the variance and covariance of the product of Z_s for 3 or 4 sets $s \in V(G_n)_{|V(H)|}$, respectively. These will be used to control the error terms in Stein's method.

Lemma 3.3. *Let $s_1, s_2, s_3 \in V(G_n)_{|V(H)|}$ be such that $\min\{|s_1 \cap s_2|, |s_1 \cap s_3|\} \geq 1$. Then the following hold:*

- (a) *If $|\bigcup_{a=1}^3 s_a| = 3|V(H)| - 2$, then $\mathbb{E}(|Z_{s_1}|Z_{s_2}Z_{s_3}) = 0$.*
- (b) *If $|\bigcup_{a=1}^3 s_a| \leq 3|V(H)| - 2$, then $\mathbb{E}|Z_{s_1}Z_{s_2}Z_{s_3}| \leq \frac{8}{c_n^{|\bigcup_{a=1}^3 s_a|-1}}$.*

Proof. We begin with the proof of case (a). First, we show that $|s_2 \cap (s_1 \cup s_3)| \leq 1$. To see this, observe

$$\begin{aligned} 3|V(H)| - 2 &= \left| \bigcup_{a=1}^3 s_a \right| = |V(H)| + |s_1 \cup s_3| - |(s_1 \cup s_3) \cap s_2| \\ &\leq |V(H)| + 2|V(H)| - 1 - |(s_1 \cup s_3) \cap s_2|, \end{aligned}$$

which implies $|s_2 \cap (s_1 \cup s_3)| \leq 1$. Therefore, $|s_2 \cap (s_1 \cup s_3)| = 1$, since $\min\{|s_1 \cap s_2|, |s_1 \cap s_3|\} \geq 1$. Let $s_2 \cap (s_1 \cup s_3) = \{j\}$, for some $j \in [|V(G_n)|]$. Then

$$\mathbb{E}(|Z_{s_1}|Z_{s_2}Z_{s_3}) = \mathbb{E}(\mathbb{E}(Z_{s_2}|X_j, \{X_i, i \in [|V(G_n)|] \setminus \{s_2\}\})|Z_{s_1}|Z_{s_3}) = 0,$$

since $\mathbb{P}(X_j = X_{q_2} \cdots = X_{q_{|V(H)|}}|X_j) = \frac{1}{c_n^{|V(H)|-1}}$, for any $q_2, q_3, \dots, q_{|V(H)|} \in V(G_n) \setminus \{j\}$ distinct. This completes the proof of (a).

Next, we prove case (b). By a direct expansion, it follows that

$$\begin{aligned} \mathbb{E}|Z_{s_1}Z_{s_2}Z_{s_3}| &\leq \mathbb{E}\mathbf{1}\{X_{=s_1}\}\mathbf{1}\{X_{=s_2}\}\mathbf{1}\{X_{=s_3}\} + \frac{4}{c_n^{3|V(H)|-3}} + \frac{\sum_{1 \leq a < b \leq 3} \mathbb{E}\mathbf{1}\{X_{=s_a}\}\mathbf{1}\{X_{=s_b}\}}{c_n^{|V(H)|-1}} \\ &= \frac{5}{c_n^{|\bigcup_{a=1}^3 s_a|-1}} + \frac{\sum_{1 \leq a < b \leq 3} \mathbb{E}\mathbf{1}\{X_{=s_a}\}\mathbf{1}\{X_{=s_b}\}}{c_n^{|V(H)|-1}}, \end{aligned} \quad (3.4)$$

since $|\bigcup_{a=1}^3 s_a| \leq 3|V(H)| - 2$. To bound the second term in the RHS above, note that

$$\left| \bigcup_{a=1}^3 s_a \right| = |s_1 \cup s_2| + |V(H)| - |(s_1 \cup s_2) \cap s_3| \leq |s_1 \cup s_2| + |V(H)| - 1, \quad (3.5)$$

since $|s_3 \cap (s_1 \cup s_2)| \geq 1$. Then Lemma 3.2(a) gives

$$\begin{aligned} \frac{1}{c_n^{|V(H)|-1}} \mathbb{E}\mathbf{1}\{X_{=s_1}\}\mathbf{1}\{X_{=s_2}\} &= \frac{1}{c_n^{|V(H)|-1}} \max \left(\frac{1}{c_n^{2|V(H)|-2}}, \frac{1}{c_n^{|s_1 \cup s_2|-1}} \right) \\ &\leq \max \left(\frac{1}{c_n^{3|V(H)|-3}}, \frac{1}{c_n^{|s_1 \cup s_2 \cup s_3|}} \right) \quad (\text{by (3.5)}) \\ &\leq \frac{1}{c_n^{|\bigcup_{a=1}^3 s_a|-1}} \quad (\text{using } |\bigcup_{a=1}^3 s_a| \leq 3|V(H)| - 2). \end{aligned}$$

Similar estimates hold for the pairs (s_1, s_3) and (s_2, s_3) as well, and, therefore, the RHS of (3.4) is bounded by $\frac{8}{c_n^{|\bigcup_{a=1}^3 s_a|-1}}$. \square

Lemma 3.4. Let $s_1, s_2, s_3, s_4 \in V(G_n) \setminus V(H)$ be such that $\min\{|s_1 \cap s_2|, |s_3 \cap s_4|\} \geq 1$. Then the following hold:

- (a) If $|\bigcup_{a=1}^4 s_a| \in \{4|V(H)| - 2, 4|V(H)| - 3\}$, then $\text{Cov}(Z_{s_1}Z_{s_2}, Z_{s_3}Z_{s_4}) = 0$.
 (b) If $|\bigcup_{a=1}^4 s_a| \leq 4|V(H)| - 4$, then $\text{Cov}(Z_{s_1}Z_{s_2}, Z_{s_3}Z_{s_4}) \lesssim \frac{1}{c_n^{|\bigcup_{a=1}^4 s_a| - 1}}$.

Proof. We begin with the proof of case (a). First, consider $|\bigcup_{a=1}^4 s_a| = 4|V(H)| - 2$, then the index sets $s_1 \cup s_2$ and $s_3 \cup s_4$ are disjoint, and so, $\text{Cov}(Z_{s_1}Z_{s_2}, Z_{s_3}Z_{s_4}) = 0$.

If $|\bigcup_{a=1}^4 s_a| = 4|V(H)| - 3$ and $|(s_1 \cup s_2) \cap (s_3 \cup s_4)| = 0$, then the index sets $s_1 \cup s_2$ and $s_3 \cup s_4$ are disjoint, and $\text{Cov}(Z_{s_1}Z_{s_2}, Z_{s_3}Z_{s_4}) = 0$. Then the assumptions $|\bigcup_{a=1}^4 s_a| = 4|V(H)| - 3$ and $\min\{|s_1 \cap s_2|, |s_3 \cap s_4|\} \geq 1$ implies, $|(s_1 \cup s_2) \cap (s_3 \cup s_4)| = 1$. In this case, we must also have $|s_1 \cup s_2| = |s_3 \cup s_4| = 2|V(H)| - 1$, which implies

$$\mathbb{E}Z_{s_1}Z_{s_2} = \mathbb{E}Z_{s_3}Z_{s_4} = 0 \quad (3.6)$$

by Lemma 3.2(a). Now, because $|(s_1 \cup s_2) \cap (s_3 \cup s_4)| = 1$, one of the sets $s_1 \cap (s_3 \cup s_4)$ and $s_2 \cap (s_3 \cup s_4)$ is non-empty. Assuming, without loss of generality, that $s_1 \cap (s_3 \cup s_4)$ is non-empty, gives

$$|s_1 \cup s_3 \cup s_4| = |s_1| + |s_3 \cup s_4| - |s_1 \cap (s_3 \cup s_4)| \leq |V(H)| + 2|V(H)| - 1 - 1 = 3|V(H)| - 2,$$

and

$$1 \leq |s_2 \cap (s_1 \cup s_3 \cup s_4)| = |V(H)| + |s_1 \cup s_3 \cup s_4| - (4|V(H)| - 3) \leq 1.$$

Denoting $s_2 \cap (s_1 \cup s_3 \cup s_4) = \{j\}$, then gives

$$\mathbb{E}Z_{s_1}Z_{s_2}Z_{s_3}Z_{s_4} = \mathbb{E}(\mathbb{E}(Z_{s_2}|X_j, \{X_i, i \in [V(G_n)] \setminus \{s_2\}\})Z_{s_1}Z_{s_3}Z_{s_4}) = 0. \quad (3.7)$$

Combining (3.6) and (3.7), $\text{Cov}(Z_{s_1}Z_{s_2}, Z_{s_3}Z_{s_4}) = 0$, whenever $|\bigcup_{a=1}^4 s_a| = 4|V(H)| - 3$. This completes the proof of (a).

Next, we prove case (b). Without loss of generality, assume $|(s_1 \cup s_2) \cap (s_3 \cup s_4)| \geq 1$, because otherwise the covariance is 0 to begin with. As in Lemma 3.3, it suffices to find bounds up to fourth joint moments of $\mathbf{1}\{X_{=s}\}$. To this end, we first claim that

$$\frac{1}{c_n^{|V(H)|-1}} \mathbb{E}\mathbf{1}\{X_{=s_1}\}\mathbf{1}\{X_{=s_2}\}\mathbf{1}\{X_{=s_3}\} \leq \frac{1}{c_n^{|\bigcup_{a=1}^4 s_a| - 1}}, \quad (3.8)$$

and a similar bound holds for all other three triples. For proving (3.8), note that $|(s_1 \cup s_2) \cap s_4| \geq |s_3 \cap s_4| \geq 1$ which gives

$$\left| \bigcup_{a=1}^4 s_a \right| = \left| \bigcup_{a=1}^3 s_a \right| + |V(H)| - \left| \left(\bigcup_{a=1}^3 s_a \right) \cap s_4 \right| \leq \left| \bigcup_{a=1}^3 s_a \right| + |V(H)| - 1. \quad (3.9)$$

Now, there are two cases:

- $|(s_1 \cup s_2) \cap s_3| \geq 1$: In this case, by (3.9),

$$\frac{1}{c_n^{|V(H)|-1}} \mathbb{E}\mathbf{1}\{X_{=s_1}\}\mathbf{1}\{X_{=s_2}\}\mathbf{1}\{X_{=s_3}\} = \frac{1}{c_n^{|\bigcup_{a=1}^3 s_a| + |V(H)| - 2}} \leq \frac{1}{c_n^{|\bigcup_{a=1}^4 s_a| - 1}}, \quad (3.10)$$

and so we have verified (3.8).

- $|(s_1 \cup s_2) \cap s_3| = 0$: In this case,

$$\frac{1}{c_n^{|V(H)|-1}} \mathbb{E}\mathbf{1}\{X_{=s_1}\}\mathbf{1}\{X_{=s_2}\}\mathbf{1}\{X_{=s_3}\} = \frac{1}{c_n^{|s_1 \cup s_2| + 2|V(H)| - 3}} \leq \frac{1}{c_n^{|\bigcup_{a=1}^4 s_a| - 1}},$$

which verifies (3.8). Note that in the last inequality we have used the bound

$$\begin{aligned} \left| \bigcup_{a=1}^4 s_a \right| &= |s_1 \cup s_2| + |s_3 \cup s_4| - |(s_1 \cup s_2) \cap (s_3 \cup s_4)| \\ &\leq |s_1 \cup s_2| + 2|V(H)| - 2, \end{aligned}$$

using $|s_3 \cup s_4| \leq 2|V(H)| - 1$, since $|s_3 \cap s_4| \geq 1$, and $|(s_1 \cup s_2) \cap (s_3 \cup s_4)| \geq 1$.

Next, proceeding to bound expectations of two tuples, observe that $|s_3 \cup s_4| - |(s_1 \cup s_2) \cap (s_3 \cup s_4)| \leq 2|V(H)| - 2$, (since $|(s_1 \cup s_2) \cap (s_3 \cup s_4)| \geq 1$, $|s_1 \cup s_2| \leq 2|V(H)| - 1$.) This implies

$$\left| \bigcup_{a=1}^4 s_a \right| = |s_1 \cup s_2| + |s_3 \cup s_4| - |(s_1 \cup s_2) \cap (s_3 \cup s_4)| \leq |s_1 \cup s_2| + 2|V(H)| - 2,$$

and so

$$\begin{aligned} \frac{1}{c_n^{2|V(H)|-2}} \mathbb{E} \mathbf{1}_{\{X=s_1\}} \mathbf{1}_{\{X=s_2\}} &\leq \frac{1}{c_n^{2|V(H)|-2}} \max \left(\frac{1}{c_n^{2|V(H)|-2}}, \frac{1}{c_n^{|s_1 \cup s_2|-1}} \right) \\ &\leq \max \left(\frac{1}{c_n^{4|V(H)|-4}}, \frac{1}{c_n^{|\bigcup_{a=1}^4 s_a|-1}} \right), \end{aligned} \quad (3.11)$$

and a similar bound applies for all the other five pairs. Thus, expanding the fourth moment and using (3.8) and (3.11) gives

$$\mathbb{E}|Z_{s_1} Z_{s_2} Z_{s_3} Z_{s_4}| \lesssim \frac{1}{c_n^{|\bigcup_{a=1}^4 s_a|-1}} + \frac{1}{c_n^{4|V(H)|-4}} \lesssim \frac{1}{c_n^{|\bigcup_{a=1}^4 s_a|-1}}. \quad (3.12)$$

Finally, by Lemma 3.2(b), if $\max(|s_1 \cup s_2|, |s_3 \cup s_4|) \geq 2|V(H)| - 1$, then $\mathbb{E} Z_{s_1} Z_{s_2} \mathbb{E} Z_{s_3} Z_{s_4} = 0$. Thus, assume that $|s_1 \cup s_2| \leq 2|V(H)| - 2$, $|s_3 \cup s_4| \leq 2|V(H)| - 2$, and so

$$|\mathbb{E} Z_{s_1} Z_{s_2} \mathbb{E} Z_{s_3} Z_{s_4}| \leq \frac{1}{c_n^{|s_1 \cup s_2| + |s_3 \cup s_4| - 2}} \leq \frac{1}{c_n^{|\bigcup_{a=1}^4 s_a|-1}}, \quad (3.13)$$

where the last inequality uses the fact that $|(s_1 \cup s_2) \cap (s_3 \cup s_4)| \geq 1$. Combining (3.12) along with (3.13) completes the proof of the lemma. \square

Proof of Theorem 1.2. Recall $Z(H, G_n)$ from (1.9). Note that (recalling (3.2)),

$$Z(H, G_n) = \frac{\sum_{s \in V(G_n) \setminus |V(H)|} \prod_{(a,b) \in E(H)} a_{s_a s_b}(G_n) Z_s}{|Aut(H)| \sqrt{\text{Var}(T(H, G_n))}} = \frac{\sum_{s \in \mathcal{A}_H^{(n)}} Z_s}{|Aut(H)| \sqrt{\text{Var}(T(H, G_n))}},$$

where $\mathcal{A}_H^{(n)}$ be the set all $s \in V(G_n) \setminus |V(H)|$ such that $\prod_{(a,b) \in E(H)} a_{s_a s_b}(G_n) = 1$. Now, for every $s_1 \in \mathcal{A}_H^{(n)}$ let

$$N_{s_1} := \{s_2 \in \mathcal{A}_H^{(n)} : |s_1 \cap s_2| \geq 1\}.$$

In other word, N_{s_1} is the subset of tuples in $\mathcal{A}_H^{(n)}$ which have at least one index common with s_1 . Then by the Stein's method based on dependency graphs (see proof of Lemma 1 in [12]), we have

$$\text{Wass}(Z(H, G_n), N(0, 1)) \lesssim_H R_1 + R_2, \quad (3.14)$$

where

$$R_1 = \left(\text{Var} \left(\frac{\sum_{s_1 \in \mathcal{A}_H^{(n)}} Z_{s_1} \sum_{s_2 \in N_{s_1}} Z_{s_2}}{\sigma_n^2} \right) \right)^{\frac{1}{2}}, \quad R_2 = \frac{\sum_{s_1 \in \mathcal{A}_H^{(n)}} \mathbb{E}|Z_{s_1}| \left(\sum_{s_2 \in N_{s_1}} Z_{s_2} \right)^2}{\sigma_n^3},$$

with $\sigma_n^2 = \text{Var}(T(H, G_n))$.

We will bound each of the terms above separately. To begin with, observe

$$\begin{aligned} \text{Var} \left(\frac{1}{\sigma_n^2} \sum_{s_1 \in \mathcal{A}_H^{(n)}} Z_{s_1} \sum_{s_2 \in N_{s_1}} Z_{s_2} \right) \\ = \frac{1}{\sigma_n^4} \sum_{s_1 \in \mathcal{A}_H^{(n)}} \sum_{s_2 \in N_{s_1}} \sum_{s_3 \in \mathcal{A}_H^{(n)}} \sum_{s_4 \in N_{s_3}} \text{Cov}(Z_{s_1} Z_{s_2}, Z_{s_3} Z_{s_4}). \end{aligned}$$

Let $\ell = |\bigcup_{a=1}^4 s_a|$ and use Lemma 3.4(a) to conclude that the above covariance vanishes unless $\ell \leq 4|V(H)| - 4$. Thus, using Lemma 3.4(b), an upper bound to the RHS above is given by

$$\begin{aligned} \frac{1}{\sigma_n^4} \sum_{\ell=|V(H)|}^{4|V(H)|-4} \frac{|V(G_n)|^\ell}{c_n^{\ell-1}} &\lesssim_H \frac{1}{\sigma_n^4} \left(\frac{|V(G_n)|^{|V(H)|}}{c_n^{|V(H)|-1}} + \frac{|V(G_n)|^{4|V(H)|-4}}{c_n^{4|V(H)|-5}} \right) \\ &\lesssim \frac{|V(G_n)|^{|V(H)|}}{c_n^{|V(H)|-1}} \times \frac{c_n^{2|V(H)|-2}}{|V(G_n)|^{2|V(H)|}} + \frac{|V(G_n)|^{4|V(H)|-4}}{c_n^{4|V(H)|-5}} \times \frac{c_n^{4|V(H)|-6}}{|V(G_n)|^{4|V(H)|-4}} \\ &= \frac{c_n^{|V(H)|-1}}{|V(G_n)|^{|V(H)|}} + \frac{1}{c_n}, \end{aligned}$$

where the last inequality uses Lemma 3.2(c). Therefore

$$R_1^2 \lesssim_H \text{Var} \left(\frac{1}{\sigma_n^2} \sum_{s_1 \in \mathcal{A}_H^{(n)}} Z_{s_1} \sum_{s_2 \in N_{s_1}} Z_{s_2} \right) \lesssim \frac{c_n^{|V(H)|-1}}{|V(G_n)|^{|V(H)|}} + \frac{1}{c_n}. \quad (3.15)$$

Proceeding to bound R_2 in (3.14), gives

$$R_2 = \frac{1}{\sigma_n^3} \sum_{s_1 \in \mathcal{A}_H^{(n)}} \mathbb{E}|Z_{s_1}| \left(\sum_{s_2 \in N_{s_1}} Z_{s_2} \right)^2 = \frac{1}{\sigma_n^3} \sum_{s_1 \in \mathcal{A}_H^{(n)}} \sum_{s_2, s_3 \in N_{s_1}} \mathbb{E}|Z_{s_1}| Z_{s_2} Z_{s_3}$$

Again let $\ell = |\bigcup_{a=1}^3 s_a|$ and use Lemma 3.3(a) to conclude that the above vanishes when $\ell = 3|V(H)| - 2$. Thus, using the bound in Lemma 3.3(b), an upper bound to the RHS above is

$$\begin{aligned} \frac{1}{\sigma_n^3} \sum_{\ell=|V(H)|}^{3|V(H)|-3} \frac{|V(G_n)|^\ell}{c_n^{\ell-1}} &\lesssim \frac{1}{\sigma_n^3} \left(\frac{|V(G_n)|^{|V(H)|}}{c_n^{|V(H)|-1}} + \frac{|V(G_n)|^{3|V(H)|-3}}{c_n^{3|V(H)|-4}} \right) \\ &\lesssim \frac{|V(G_n)|^{|V(H)|}}{c_n^{|V(H)|-1}} \times \frac{c_n^{\frac{3}{2}(|V(H)|-1)}}{|V(G_n)|^{\frac{3}{2}|V(H)|}} + \frac{|V(G_n)|^{3|V(H)|-3}}{c_n^{3|V(H)|-4}} \times \frac{c_n^{\frac{3}{2}(2|V(H)|-3)}}{|V(G_n)|^{3|V(H)|-3}} \\ &= \left(\frac{c_n^{|V(H)|-1}}{|V(G_n)|^{|V(H)|}} \right)^{\frac{1}{2}} + \left(\frac{1}{c_n} \right)^{\frac{1}{2}}, \end{aligned}$$

where again the last inequality uses Lemma 3.2(c). Therefore,

$$R_2 = \frac{1}{\sigma_n^3} \sum_{s_1 \in V(G_n) \setminus V(H)} \mathbb{E}|Z_{s_1}| \left(\sum_{s_2 \in N_{s_1}} Z_{s_2} \right)^2 \lesssim \left(\frac{c_n^{|V(H)|-1}}{|V(G_n)|^{|V(H)|}} \right)^{\frac{1}{2}} + \left(\frac{1}{c_n} \right)^{\frac{1}{2}}. \quad (3.16)$$

Combining (3.14) with (3.15) and (3.16) completes the proof of (1.10).

To see that the error term in the RHS of (1.10) goes to zero, first note that $\mathbb{E}t(H, G_n) = \frac{1}{|Aut(H)|} \cdot \frac{\text{hom}_{\text{inj}}(H, G_n)}{c_n^{|V(H)|-1}}$ (recall (1.7)). Then using $\frac{1}{|V(G_n)|^{|V(H)|}} \text{hom}_{\text{inj}}(H, G_n) \rightarrow t(H, W)$, as G_n converges

to the graphon W , it follows that the first term in the RHS of [Theorem 1.2](#) goes to zero whenever $\mathbb{E}T(H, G_n) \rightarrow \infty$. Therefore, $Z(H, G_n) \xrightarrow{D} N(0, 1)$, whenever $\mathbb{E}T(H, G_n) \rightarrow \infty$ and $c_n \rightarrow \infty$. \square

As noted in [Remark 1.2](#), the asymptotic normality of $Z(K_2, H)$ holds for all graph sequences $\{G_n\}_{n \geq 1}$, as long as $\mathbb{E}T(K_2, H) \rightarrow \infty$ and $c_n \rightarrow \infty$. However, we cannot expect the normality of $Z(H, G_n)$, for general graphs H , to extend beyond dense graphs, without further assumptions, as shown below:

Example 4. For $n \geq 1$, denote by $\mathcal{D}_n = (V(\mathcal{D}_n), E(\mathcal{D}_n))$ the n -pyramid: $V(\mathcal{D}_n) = \{a, b, c_1, c_2, \dots, c_n\}$ and

$$E(\mathcal{D}_n) = \{(a, b), (a, c_1), (a, c_2), \dots, (a, c_n), (b, c_1), (b, c_2), \dots, (b, c_n)\}.$$

In other words, the n -pyramid is the complete 3-partite graph $K_{1,1,n}$. (Note that the 1-pyramid is the triangle ($\mathcal{D}_1 = K_3$) and the 2-pyramid \mathcal{D}_2 is the 4-cycle with a diagonal.) Now, let G_n be the disjoint union of the n -pyramid \mathcal{D}_n and the complete bipartite graph $K_{n,n}$ and $H = K_3$ the triangle. Choose $c_n \rightarrow \infty$ such that $c_n = o(\sqrt{n})$. In this case, $\mathbb{E}T(K_3, G_n) = \frac{n}{c_n^2} \rightarrow \infty$, but $\lim_{n \rightarrow \infty} t(K_3, G_n) = 0$, therefore, [Theorem 1.2](#) does not apply. Moreover, as every triangle in G_n must pass through the base vertices a, b of the pyramid \mathcal{D}_n ,

$$\mathbb{P}(T(K_3, G_n) > 0) \leq \mathbb{P}(X_a = X_b) = \frac{1}{c_n} \rightarrow 0.$$

This implies, $T(K_3, G_n) \xrightarrow{P} 0$, and hence, $Z(K_3, G_n) \xrightarrow{P} 0$, that is, $Z(K_3, G_n)$ does not converge to a non-degenerate limiting distribution.

4. Limiting distribution for fixed number of colors

In this section we derive the limiting distribution for the number of monochromatic subgraphs when the number of colors is fixed. The proof of [Theorem 1.3](#) is given in [Section 4.1](#). Examples are discussed in [Section 4.2](#).

4.1. Proof of [Theorem 1.3](#)

We begin with the following observation:

Observation 4.1. For $\mathbf{s} \in V(G_n)^{|V(H)|}$, let $Z_{\mathbf{s}} = \mathbf{1}\{X_{\mathbf{s}}\} - \frac{1}{c^{|V(H)|-1}}$. Then

$$Z_{\mathbf{s}} = \sum_{a=1}^c \sum_{\substack{J \subseteq V(H) \\ |J| \geq 2}} \frac{1}{c^{|V(H)|-|J|}} \prod_{j \in J} \left(\mathbf{1}\{X_{s_j} = a\} - \frac{1}{c} \right). \quad (4.1)$$

Proof. To begin with note that

$$\frac{1}{c^{|V(H)|-1}} \sum_{a=1}^c \sum_{\substack{J \subseteq V(H) \\ |J|=1}} \prod_{j \in J} \left(\mathbf{1}\{X_{s_j} = a\} - \frac{1}{c} \right) = \frac{1}{c^{|V(H)|-1}} \sum_{a=1}^c \sum_{v=1}^{|V(H)|} \left(\mathbf{1}\{X_v = a\} - \frac{1}{c} \right) = 0,$$

since, for every $v \in V(G_n)$, $\sum_{a=1}^c (\mathbf{1}\{X_v = a\} - \frac{1}{c}) = 0$. Therefore, denoting the RHS of [\(4.1\)](#) by $\tilde{Z}_{\mathbf{s}}$, we get

$$\tilde{Z}_{\mathbf{s}} = \sum_{a=1}^c \sum_{J \subseteq V(H)} \frac{1}{c^{|V(H)|-|J|}} \prod_{j \in J} \left(\mathbf{1}\{X_{s_j} = a\} - \frac{1}{c} \right) - \frac{1}{c^{|V(H)|-1}}, \quad (4.2)$$

where the $\frac{1}{c^{|V(H)|-1}}$ adjustment cancels with the term corresponding to taking $J = \emptyset$ in the sum. Here, we use the convention that $\prod_{j \in J} (\mathbf{1}\{X_{s_j} = a\} - \frac{1}{c}) = 1$ if $J = \emptyset$. Now, expanding the product $\prod_{j \in J} (\mathbf{1}\{X_{s_j} = a\} - \frac{1}{c})$ in the RHS above gives,

$$\begin{aligned} \sum_{J \subseteq V(H)} \frac{1}{c^{|V(H)|-|J|}} \prod_{j \in J} \left(\mathbf{1}\{X_{s_j} = a\} - \frac{1}{c} \right) &= \sum_{J \subseteq V(H)} \frac{1}{c^{|V(H)|-|J|}} \sum_{J' \subseteq J} \frac{(-1)^{|J|-|J'|}}{c^{|J|-|J'|}} \prod_{j \in J'} \mathbf{1}\{X_{s_j} = a\} \\ &= \sum_{J \subseteq V(H)} \sum_{J' \subseteq J} \frac{(-1)^{|J|-|J'|}}{c^{|V(H)|-|J'|}} \prod_{j \in J'} \mathbf{1}\{X_{s_j} = a\} \\ &= \sum_{s=0}^{|V(H)|} \sum_{\substack{J \subseteq V(H) \\ |J|=s}} \sum_{J' \subseteq J} \frac{(-1)^{s-|J'|}}{c^{|V(H)|-|J'|}} X_a(J'), \end{aligned} \quad (4.3)$$

where $X_a(J') = \prod_{j \in J'} \mathbf{1}\{X_{s_j} = a\}$. The RHS of (4.3) can be rewritten as the weighted sum over subsets $J' \subseteq V(H)$. Note that the coefficient of $X(J')$ in the above sum is

$$\sum_{s=|J'|}^{|V(H)|} \binom{|V(H)|-|J'|}{s-|J'|} \frac{(-1)^{s-|J'|}}{c^{|V(H)|-|J'|}} = \frac{1}{c^{|V(H)|-|J'|}} \sum_{t=0}^{|V(H)|-|J'|} \binom{|V(H)|-|J'|}{t} (-1)^t = 0,$$

whenever $|V(H)| \neq |J'|$. This follows by noting that there is $\binom{|V(H)|-|J'|}{s-|J'|}$ ways of completing the given set J' to a s -element subset of $|V(H)|$. Therefore, the only term in (4.3) which is non-zero corresponds to taking $J' = V(H)$, and (4.2) simplifies to

$$\begin{aligned} \tilde{Z}_s &= \sum_{a=1}^c X_a(V(H)) - \frac{1}{c^{|V(H)|-1}} = \sum_{a=1}^c \prod_{j=1}^{|V(H)|} \mathbf{1}\{X_{s_j} = a\} - \frac{1}{c^{|V(H)|-1}} \\ &= \mathbf{1}\{X=s\} - \frac{1}{c^{|V(H)|-1}} = Z_s, \end{aligned}$$

as required. \square

Using this observation, $T(H, G_n)$ can be written as a polynomial in the i.i.d. color vectors $\{(\mathbf{1}\{X_v = a\})_{a \in [c]} : v \in V(G_n)\}$.

$$\begin{aligned} T(H, G_n) - \mathbb{E}(T(H, G_n)) &= \sum_{\mathbf{s} \in V(G_n) \setminus V(H)} M_{G_n}(\mathbf{s}, H) Z_{\mathbf{s}} \quad (\text{recall (3.2)}) \\ &= \sum_{\mathbf{s} \in V(G_n) \setminus V(H)} M_{G_n}(\mathbf{s}, H) \sum_{a=1}^c \sum_{\substack{J \subseteq V(H) \\ |J| \geq 2}} \frac{1}{c^{|V(H)|-|J|}} \prod_{j \in J} \left(\mathbf{1}\{X_{s_j} = a\} - \frac{1}{c} \right) \\ &= \sum_{\substack{J \subseteq V(H) \\ |J| \geq 2}} T_J(H, G_n), \end{aligned} \quad (4.4)$$

where

$$T_J(H, G_n) := \sum_{a=1}^c \sum_{\mathbf{s} \in V(G_n) \setminus V(H)} M_{G_n}(\mathbf{s}, H) \frac{1}{c^{|V(H)|-|J|}} \prod_{j \in J} \left(\mathbf{1}\{X_{s_j} = a\} - \frac{1}{c} \right). \quad (4.5)$$

Lemma 4.1. For every $J \subseteq V(H)$ such that $|J| \geq 3$, $T_J(H, G_n) = o_P(|V(G_n)|^{|V(H)|-1})$.

Proof. Fix $J \subseteq V(H)$ such that $|J| \geq 3$. To begin with note that

$$\mathbb{E}T_J(H, G_n) = \sum_{a=1}^c \sum_{\mathbf{s} \in V(G_n) \setminus V(H)} M_{G_n}(\mathbf{s}, H) \frac{1}{c^{|V(H)|-|J|}} \mathbb{E} \left(\prod_{j \in J} \mathbf{1}\{X_{s_j} = a\} - \frac{1}{c} \right) = 0, \quad (4.6)$$

since

$$\mathbb{E} \left(\prod_{j \in J} \mathbf{1}\{X_{s_j} = a\} - \frac{1}{c} \right) = \prod_{j \in J} \mathbb{E} \left(\mathbf{1}\{X_{s_j} = a\} - \frac{1}{c} \right) = 0,$$

since $\{\mathbf{1}\{X_{s_j} = a\} - \frac{1}{c} : j \in J\}$ is a collection of independent random variables.

The second moment of $T_J(H, G_n)$ equals

$$\sum_{a, a' \in [c]} \sum_{\mathbf{s}, \mathbf{s}' \in V(G_n) \setminus V(H)} \frac{M_{G_n}(\mathbf{s}, H) M_{G_n}(\mathbf{s}', H)}{c^{2|V(H)|-2|J|}} \mathbb{E} \prod_{j \in J} \left(\mathbf{1}\{X_{s_j} = a\} - \frac{1}{c} \right) \left(\mathbf{1}\{X_{s'_j} = a'\} - \frac{1}{c} \right). \quad (4.7)$$

Now, if there exists $s_0 \in \{s_j : j \in J\} \setminus \{s'_j : j \in J\}$, then

$$\begin{aligned} & \mathbb{E} \prod_{j \in J} \left(\mathbf{1}\{X_{s_j} = a\} - \frac{1}{c} \right) \left(\mathbf{1}\{X_{s'_j} = a'\} - \frac{1}{c} \right) \\ &= \mathbb{E} \prod_{j \in J, s_j \neq s_0} \left(\mathbf{1}\{X_{s_j} = a\} - \frac{1}{c} \right) \prod_{j \in J} \left(\mathbf{1}\{X_{s'_j} = a'\} - \frac{1}{c} \right) \mathbb{E} \left(\mathbf{1}\{X_{s_0} = a\} - \frac{1}{c} \right) = 0. \end{aligned}$$

Similarly, the expectation vanishes if $s_0 \in \{s'_j : j \in J\} \setminus \{s_j : j \in J\}$.

Now, consider $\mathbf{s}, \mathbf{s}' \in V(G_n) \setminus V(H)$ such that $\{s_j : j \in J\} = \{s'_j : j \in J\}$. Define

$$W_J(a, a') := \mathbb{E} \prod_{j \in J} \left(\mathbf{1}\{X_{s_j} = a\} - \frac{1}{c} \right) \left(\mathbf{1}\{X_{s'_j} = a'\} - \frac{1}{c} \right).$$

Note that $|W_J(a, a')| \leq 1$, for all $a, a' \in [c]$. Therefore, (4.7) gives

$$\begin{aligned} \mathbb{E}T_J(H, G_n)^2 &= \sum_{a, a' \in [c]} \sum_{\substack{\mathbf{s}, \mathbf{s}' \in V(G_n) \setminus V(H) \\ \{s_j : j \in J\} = \{s'_j : j \in J\}}} M_{G_n}(\mathbf{s}, H) M_{G_n}(\mathbf{s}', H) \frac{W_J(a, a')}{c^{2|V(H)|-2|J|}} \\ &\lesssim_c \sum_{\substack{\mathbf{s}, \mathbf{s}' \in V(G_n) \setminus V(H) \\ \{s_j : j \in J\} = \{s'_j : j \in J\}}} M_{G_n}(\mathbf{s}, H) M_{G_n}(\mathbf{s}', H) \\ &= O(|V(G_n)|^{2|V(H)|-|J|}) = o(|V(G_n)|^{2|V(H)|-2}), \end{aligned} \quad (4.8)$$

whenever $|J| \geq 3$.

Combining (4.6) and (4.8) it follows that $T_J(H, G_n) = o_P(|V(G_n)|^{|V(H)|-1})$, whenever $|J| \geq 3$. \square

Definition 4.1. Let H be a labeled finite simple graph. Then, for $1 \leq u \neq v \leq |V(H)|$ and $1 \leq i \neq j \leq |V(G_n)|$, define $M_{u,v}(i, j, H, G_n)$ as the number of injective homomorphism $\phi : V(H) \rightarrow V(G_n)$ such that $\phi(u) = i$ and $\phi(v) = j$. More formally,

$$\begin{aligned} & M_{u,v}(i, j, H, G_n) \\ &= a_{ij,uv}^+(G_n) \sum_{\substack{\mathbf{s} \setminus \{s_u, s_v\} \\ \mathbf{s} \in V(G_n) \setminus V(H)}} \prod_{x \in N_H(u) \setminus \{v\}} a_{is_x}(G_n) \prod_{y \in N_H(v) \setminus \{u\}} a_{js_y}(G_n) \prod_{(x,y) \in E(H \setminus \{u,v\})} a_{s_x s_y}(G_n), \end{aligned}$$

with $a_{ij,uv}^+(G_n) = a_{ij}(G_n)$ if $(u, v) \in E(H)$ and 1 otherwise, and the sum is over indices $\mathbf{s} \setminus \{s_u, s_v\}$, with $\mathbf{s} \in V(G_n) \setminus V(H)$, which are distinct and belong to $[|V(G_n)|] \setminus \{i, j\}$. Note that $M_{u,v}(i, j, H, G_n)$ is,

in general, not symmetric in i, j , but satisfies $M_{v,u}(i, j, H, G_n) = M_{u,v}(j, i, H, G_n)$.⁷ Finally, define the symmetric scaled 2-point homomorphism matrix as $((\bar{\mathbf{B}}_H(G_n))_{ij})_{i,j \in [V(G_n)]}$ with

$$\bar{\mathbf{B}}_H(G_n)_{ij} := \frac{1}{2|Aut(H)| \cdot |V(G_n)|^{|V(H)|-1}} \cdot \sum_{1 \leq u \neq v \leq |V(H)|} M_{u,v}(i, j, H, G_n), \quad (4.9)$$

for $1 \leq i \neq j \leq |V(G_n)|$.

The following lemma shows that $\Gamma(H, G_n)$ is a sum of c quadratic forms in terms of the scaled 2-point homomorphism matrix, up to $o_P(1)$ terms.

Lemma 4.2. Recall $\Gamma(H, G_n)$ from (1.12) and define,

$$\Gamma_2(H, G_n) := \frac{1}{c^{|V(H)|-2}} \sum_{a=1}^c \sum_{1 \leq i \neq j \leq |V(G_n)|} \bar{\mathbf{B}}_H(G_n)_{ij} \left(\mathbf{1}\{X_i = a\} - \frac{1}{c} \right) \left(\mathbf{1}\{X_j = a\} - \frac{1}{c} \right),$$

where $\bar{\mathbf{B}}_H(G_n)$ is the 2-point homomorphism matrix as defined in (4.9). Then, $\Gamma(H, G_n) = \Gamma_2(H, G_n) + o_P(1)$.

Proof. Recalling (4.5), note that

$$\begin{aligned} & c^{|V(H)|-2} \sum_{a=1}^c \sum_{\substack{J \subseteq V(H) \\ |J|=2}} T_J(H, G_n) \\ &= \sum_{a=1}^c \sum_{1 \leq u < v \leq |V(H)|} \sum_{1 \leq s_u \neq s_v \leq |V(G_n)|} \frac{M_{u,v}(s_u, s_v, H, G_n)}{|Aut(H)|} \left(\mathbf{1}\{X_{s_u} = a\} - \frac{1}{c} \right) \left(\mathbf{1}\{X_{s_v} = a\} - \frac{1}{c} \right), \\ &= \sum_{a=1}^c \sum_{1 \leq u < v \leq |V(H)|} \sum_{1 \leq i \neq j \leq |V(G_n)|} \frac{M_{u,v}(i, j, H, G_n)}{|Aut(H)|} \left(\mathbf{1}\{X_i = a\} - \frac{1}{c} \right) \left(\mathbf{1}\{X_j = a\} - \frac{1}{c} \right), \\ &= |V(G_n)|^{|V(H)|-1} \cdot \sum_{a=1}^c \sum_{1 \leq i \neq j \leq |V(G_n)|} \mathbf{B}_H(G_n)_{ij} \left(\mathbf{1}\{X_i = a\} - \frac{1}{c} \right) \left(\mathbf{1}\{X_j = a\} - \frac{1}{c} \right), \end{aligned}$$

where $\mathbf{B}_H(G_n) = ((\mathbf{B}_H(G_n))_{ij})_{i,j \in [V(G_n)]}$ is a matrix with

$$\mathbf{B}_H(G_n)_{ij} = \frac{1}{|V(G_n)|^{|V(H)|-1}} \cdot \frac{1}{|Aut(H)|} \sum_{1 \leq u < v \leq |V(H)|} M_{u,v}(i, j, H, G_n).$$

Now, from (4.9) it is easy to see that $\bar{\mathbf{B}}_H(G_n)_{ij} = \frac{\mathbf{B}_H(G_n)_{ij} + \mathbf{B}_H(G_n)_{ji}}{2}$, which along with Lemma 4.1 gives the desired conclusion. \square

Next, define the analogous random variable for $\Gamma_2(H, G_n)$, where the centered color vectors $\{\mathbf{R}_v : v \in V(G_n)\}$, where $\mathbf{R}_v = (\mathbf{1}\{X_v = a\} - \frac{1}{c})_{a \in [c]}$, are replaced by a collection of i.i.d. Gaussian vectors with the same mean and covariance structure. More formally,

$$Q_2(H, G_n) := \frac{1}{c^{|V(H)|-2}} \sum_{a=1}^c \sum_{1 \leq i \neq j \leq |V(G_n)|} \bar{\mathbf{B}}_H(G_n)_{ij} \hat{U}_{i,a} \hat{U}_{j,a}. \quad (4.10)$$

with $\hat{U}_{v,a} = U_{v,a} - \bar{U}_{v,a}$, where $\{U_{v,a} : v \in V(G_n), a \in [c]\}$ are i.i.d. Gaussians with mean 0 and variance $1/c$ random variables and $\bar{U}_{v,a} = \frac{1}{c} \sum_{a=1}^c U_{v,a}$. Note that for each $v \in V(G_n)$ the random vector

⁷ For example, when $H = K_{1,2}$ is the 2-star with the central vertex labeled 1, then $M_{1,2}(i, j, K_{1,2}, G_n) = M_{1,3}(i, j, K_{1,2}, G_n) = a_{ij}(G_n)(d_{G_n}(i) - a_{ij}(G_n))$, where $d_{G_n}(i)$ is the degree of the vertex i in G_n , and $M_{2,3}(i, j, K_{1,2}, G_n) = \sum_{k \neq \{i,j\}} a_{ik}(G_n)a_{jk}(G_n)$, the number of common neighbors of i, j .

$\hat{\mathbf{U}}_v := (\hat{U}_{v,1}, \hat{U}_{v,2}, \dots, \hat{U}_{v,c})$ has mean 0 and the same covariance matrix as \mathbf{R}_v . Also, $\{\hat{\mathbf{U}}_v, v \in V(G_n)\}$ are independent and identically distributed random vectors. Finally, define

$$\Delta_2(H, G_n) := \frac{Q_2(H, G_n)}{|V(G_n)|^{|V(H)|-1}}. \quad (4.11)$$

The next lemma shows that the moments of $\Gamma_2(H, G_n)$ and $\Delta_2(H, G_n)$ are asymptotically close.

Lemma 4.3. For every integer $r \geq 1$,

$$\lim_{n \rightarrow \infty} \left\{ \mathbb{E} \Gamma_2(H, G_n)^r - \mathbb{E} \Delta_2(H, G_n)^r \right\} = 0. \quad (4.12)$$

Proof. Fix an integer $r \geq 1$. Using the bound $|V(G_n)| \max_{i,j \in V(G_n)} \bar{\mathbf{B}}_H(G_n)_{ij} \lesssim_H 1$, a direct expansion gives

$$\begin{aligned} & c^{r|V(H)|-2r} \left| \mathbb{E} \Gamma_2(H, G_n)^r - \mathbb{E} \Delta_2(H, G_n)^r \right| \\ & \lesssim_H \frac{1}{|V(G_n)|^r} \sum_{a_1, \dots, a_r \in [c]} \sum_{\substack{1 \leq i_1 \neq j_1 \leq |V(G_n)| \\ \vdots \\ 1 \leq i_r \neq j_r \leq |V(G_n)|}} \left| \mathbb{E} \prod_{s=1}^r \left(\mathbf{1}\{X_{i_s} = a_s\} - \frac{1}{c} \right) \left(\mathbf{1}\{X_{j_s} = a_s\} - \frac{1}{c} \right) \right. \\ & \quad \left. - \mathbb{E} \prod_{s=1}^r \hat{U}_{i_s, a_s} \hat{U}_{j_s, a_s} \right|. \end{aligned}$$

Now, fix an index set $J := \{i_1, \dots, i_r, j_1, \dots, j_r\}$, where $1 \leq i_s \neq j_s \leq |V(G_n)|$, for $s \in [r]$. If an index in J appears exactly once, then using $\mathbb{E} \left(\mathbf{1}\{X_i = a\} - \frac{1}{c} \right) = \mathbb{E} \hat{U}_{i,a} = 0$, for every $i \in V(G_n)$ and $a \in [c]$, it is easy to see that both the moments inside the absolute value vanish. Therefore, we can assume that every index in J appears at least twice. Moreover, as the total number of terms with at most $r-1$ distinct indices from J is bounded above by $|V(G_n)|^{r-1}$, it suffices to consider the terms where the number of distinct indices from J is exactly r , up to a $o(1)$ -term. But in this case every index in J appears exactly twice, and to prove (4.12) it suffices to show for any such index set $(i_1, j_1), \dots, (i_r, j_r)$ we have

$$\left| \mathbb{E} \prod_{s=1}^r \left(\mathbf{1}\{X_{i_s} = a_s\} - \frac{1}{c} \right) \left(\mathbf{1}\{X_{j_s} = a_s\} - \frac{1}{c} \right) - \mathbb{E} \prod_{s=1}^r \hat{U}_{i_s, a_s} \hat{U}_{j_s, a_s} \right| = 0.$$

Indeed, in this case both the moments factorize over the distinct indices, and to show equality of moments it suffices to check that for all $a, b \in [c]$ we have

$$\mathbb{E} \left(\mathbf{1}\{X_i = a\} - \frac{1}{c} \right) \left(\mathbf{1}\{X_i = b\} - \frac{1}{c} \right) = \mathbb{E} \hat{U}_{i,a} \hat{U}_{i,b}.$$

This follows by noting that both sides equal $\frac{1}{c} \left(1 - \frac{1}{c} \right)$ if $a = b$, and $-\frac{1}{c^2}$ otherwise. \square

From Lemmas 4.2 and 4.3, to derive the limiting distribution of $\Gamma(H, G_n)$ it suffices to derive the limiting distribution of $\Delta_2(H, G_n)$, which is the sum of c -quadratic forms in $\bar{\mathbf{B}}_H(G_n)$. To this end, we need to understand the spectrum of the matrix $\bar{\mathbf{B}}_H(G_n)$. We begin by defining the notion of cycles formed by H , which arise in the analysis of the power-sum of the eigenvalues of $\bar{\mathbf{B}}_H(G_n)$.

Definition 4.2. Fix an integer $g \geq 2$, and let H_1, H_2, \dots, H_g be g isomorphic copies of H , where the image of the vertex $z \in V(H)$ in H_a will be denoted by $z^{(a)}$, for $a \in [g]$. Then fixing indices $J := \{(u_a, v_a) : 1 \leq u_a \neq v_a \leq |V(H)|, a \in [g]\}$, define the r -cycle of H with pivots at J as the graph obtained by the union of H_1, H_2, \dots, H_g , where the vertex $v_a^{(a)} \in V(H_a)$ identified with the vertex $u_{a+1}^{(a+1)} \in V(H_{a+1})$, for $a \in [g]$, with $u_{g+1}^{(g+1)} := u_1^{(1)}$ and $H_{g+1} = H_1$. Denote this graph by $H^{(g)}(J)$. From

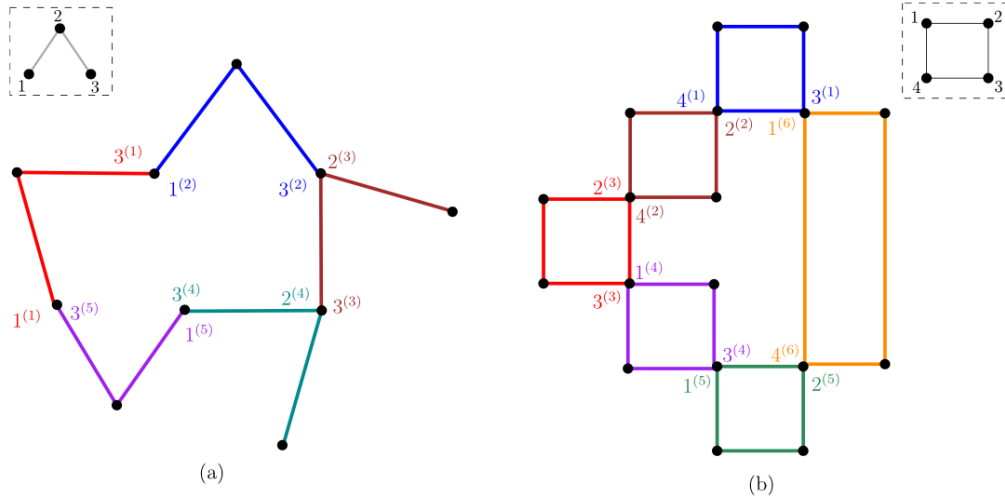


Fig. 1. A 5-cycle of $K_{1,2}$ with pivots $\{(1, 3), (1, 3), (2, 3), (2, 3), (1, 3)\}$, (b) a 6-cycle of C_4 with pivots $\{(3, 4), (2, 4), (2, 3), (1, 3), (1, 2), (4, 1)\}$.

Definition 1.2, it is easy to see that

$$t(H^{(g)}(J), W) = \int_{[0,1]^g} \prod_{a=1}^g t_{u_a, v_a}(x_a, x_{a+1}, H, W) \prod_{a=1}^g dx_a. \quad (4.13)$$

Fig. 1 shows a 5-cycle of $K_{1,2}$ and a 6-cycle of C_4 , and the associated pivots.

Equipped with the above definitions and recalling the function W_H from (1.13), we proceed to prove the convergence of the spectrum of $\bar{B}_H(G_n)$.

Lemma 4.4. Let $\{\lambda_1(\bar{B}_H(G_n)), \lambda_2(\bar{B}_H(G_n)), \dots, \lambda_{|V(G_n)|}(\bar{B}_H(G_n))\}$ be the multi-set of eigenvalues of $\bar{B}_H(G_n)$. Then, for every $g \geq 2$, $\lim_{n \rightarrow \infty} \sum_{r=1}^{|V(G_n)|} \lambda_r(\bar{B}_H(G_n))^g = \sum_{i=1}^{\infty} \lambda_i(H, W)^g$, where $\{\lambda_1(H, W), \lambda_2(H, W), \dots\}$ is the multi-set of eigenvalues of W_H . Moreover, the assumption $t(H, W) > 0$ ensures that the spectrum of W_H is non-trivial, that is, W_H has at least one (non-zero) eigenvalue.

Proof. Fix $g \geq 2$. Define $K_g := \frac{1}{2^g |Aut(H)|^g}$. Then

$$\begin{aligned} & \sum_{r=1}^{|V(G_n)|} \lambda_r(\bar{B}_H(G_n))^g \\ &= \text{tr}(\bar{B}_H(G_n)^g) \\ &= \sum_{j \in V(G_n)^g} \prod_{a=1}^g \bar{B}_H(G_n)_{j_a j_{a+1}} \quad (\text{where } j_{g+1} = j_1) \\ &= K_g \cdot \frac{1}{|V(G_n)|^{g|V(H)|-g}} \sum_{j \in V(G_n)^g} \prod_{a=1}^g \sum_{1 \leq u \neq v \leq |V(H)|} M_{u,v}(j_a, j_{a+1}, H, G_n) \quad (\text{recall (4.9)}) \\ &= K_g \cdot \frac{1}{|V(G_n)|^{g|V(H)|-g}} \sum_{1 \leq u_1 \neq v_1 \leq |V(H)|} \cdots \sum_{1 \leq u_g \neq v_g \leq |V(H)|} \sum_{j \in V(G_n)^g} \prod_{a=1}^g M_{u_a, v_a}(j_a, j_{a+1}, H, G_n). \end{aligned}$$

Now, fix $J = \{(u_a, v_a) : 1 \leq u_a \neq v_a \leq |V(H)|, a \in [g]\}$. Recalling Definition 4.1, note that $M_{u_a, v_a}(j_a, j_{a+1}, H, G_n)$ is the sum over $\mathbf{s} \setminus \{s_{u_a}, s_{v_a}\}$, where $\mathbf{s} \in V(G_n)_{|V(H)|}$, that is, the sum ranges over

$|V(H)| - 2$ indices taking values in $\{1, 2, \dots, |V(G_n)|\}$. This implies, the product

$$\sum_{j \in V(G_n)^g} \prod_{a=1}^g M_{u_a, v_a}(j_a, j_{a+1}, H, G_n), \quad (4.14)$$

can be expanded to obtain a sum over at most $g + g(|V(H)| - 2) = g|V(H)| - g$ indices taking values in $\{1, 2, \dots, |V(G_n)|\}$. However, if any two of the indices are the same, then the corresponding term in the sum is $o(|V(G_n)|^{g|V(H)| - g})$. (Here, we bound each $a_{ij}(G_n)$ by 1 and the sum over each distinct index by $|V(G_n)|$.) This shows that the leading term in (4.14) is a sum over $g|V(H)| - g$ distinct indices ranging in $V(G_n)$, which counts the number of injective homomorphisms of $H^{(g)}(J)$ in G_n , where $H^{(g)}(J)$ is the g -cycle of H with pivots at J , as in Definition 4.2. Therefore,

$$\frac{1}{|V(G_n)|^{g|V(H)| - g}} \sum_{j \in V(G_n)^g} \prod_{a=1}^g M_{u_a, v_a}(j_a, j_{a+1}, H, G_n) = t(H^{(g)}(J), G_n) + o(1) \rightarrow t(H^{(g)}(J), W).$$

Next, recall that $\lambda_1(H, W), \lambda_2(H, W), \dots$ are the eigenvalues of the function W_H , as defined in Theorem 1.3. Then by the spectral theorem [24, Section 7.5],

$$\begin{aligned} \sum_{r=1}^{\infty} \lambda_r^g(H, W) &= t(C_g, W_H) \\ &= \int_{[0,1]^g} \prod_{a=1}^g W_H(x_a, x_{a+1}) \prod_{a=1}^g dx_a \quad (x_{g+1} := x_1) \\ &= K_g \int_{[0,1]^g} \prod_{a=1}^g \sum_{1 \leq u \neq v \leq |V(H)|} t_{u,v}(x_a, x_{a+1}, H, W) \prod_{a=1}^g dx_a \quad (\text{by (1.13)}) \\ &= K_g \sum_{1 \leq u_1 \neq v_1 \leq |V(H)|} \cdots \sum_{1 \leq u_g \neq v_g \leq |V(H)|} \int_{[0,1]^g} \prod_{a=1}^g t_{u_a, v_a}(x_a, x_{a+1}, H, W) \prod_{a=1}^g dx_a \\ &= K_g \sum_{1 \leq u_1 \neq v_1 \leq |V(H)|} \cdots \sum_{1 \leq u_g \neq v_g \leq |V(H)|} t(H^{(g)}(J), W), \end{aligned} \quad (4.15)$$

where the last step uses (4.13). Therefore, (4.15) implies that the g th power sum of the eigenvalues of $\bar{B}_H(G_n)$ converge to the g th power sum of eigenvalues of W_H , for every $g \geq 2$.

Finally, note that for $g = 2$ and any set of pivots of the form $J = \{(a, b), (b, a)\}$, where $1 \leq a \neq b \leq |V(H)|$, $H^{(2)}(J) = H_{(a,b)}^2$ (recall Definition 3.1) and by Lemma 3.1, $t(H^{(2)}(J), W) > 0$. Therefore, $\sum_{r=1}^{\infty} \lambda_r^2(H, W) > 0$, which implies that the spectrum of W_H is non-trivial. \square

Having established the convergence of the spectrum of $\bar{B}_H(G_n)$, it remains to derive the asymptotic distribution of $\Delta_2(H, G_n)$, and hence $\Gamma(H, G_n)$. This follows from the lemma below, which can be easily proved by computing the moment generating function of $\Delta_2(H, G_n)$ using the spectral decomposition, as in [7, Lemma 7.3].

Lemma 4.5 ([7, Lemma 7.3]). *Let $\mathbf{Q}_n = ((Q_n(i, j)))_{i, j \in [|V(G_n)|]}$ be a sequence of symmetric $|V(G_n)| \times |V(G_n)|$ matrices with zeros on the diagonal. If there exist constants $\lambda_1, \lambda_2, \dots$ such that $\lim_{n \rightarrow \infty} \text{tr}(\mathbf{Q}_n^s) = \sum_{r=1}^{\infty} \lambda_r^s < \infty$, for every $s \geq 2$, then*

$$\sum_{a=1}^c \sum_{1 \leq i \neq j \leq |V(G_n)|} Q_n(i, j) \hat{U}_{i,a} \hat{U}_{j,a} \xrightarrow{D} \frac{1}{c} \sum_{r=1}^{\infty} \lambda_r \eta_r,$$

where $\{\eta_r\}_{r \geq 1}$ is a collection of i.i.d. $\chi_{(c-1)}^2 - (c-1)$ random variables. \square

4.2. Examples

To begin with, we consider monochromatic edges, that is, $H = K_2$. In this case, the 2-point homomorphism matrix is just the scaled adjacency matrix of G_n , and we re-derive [7, Theorem 1.4].

Example 5 (Monochromatic Edges). Let G_n be a sequence of graphs converging to the graphon W and $H = K_2$. Then $|Aut(H)| = 2$ and $W_{K_2}(x, y) = \frac{1}{2}W(x, y)$, and $\lambda_r(W_{K_2}) = \frac{1}{2}\lambda_r(W)$, where $\lambda_1(W), \lambda_2(W), \dots$ are the eigenvalues of the operator $T_W : L_2[0, 1] \rightarrow L_2[0, 1]$, defined as $(T_W f)(x) = \int_0^1 W(x, y)f(y)dy$. Then Theorem 1.3 shows

$$\Gamma(K_2, G_n) \xrightarrow{D} \frac{1}{2c} \sum_{r=1}^{\infty} \lambda_r(W) \eta_r,$$

where $\{\eta_r\}_{r \in \mathbb{N}}$ are independent $\chi_{(c-1)}^2 - (c-1)$ random variables, as in [7, Theorem 1.4].

As before, Theorem 1.3 applies to convergent sequence of dense random graphs, when the limit in (1.5) holds in probability.

Example 6 (Erdős–Rényi Random Graph). Let $G_n \sim G(n, p)$ be the Erdős–Rényi random graph and H be any finite simple graph. In this case, G_n converges to the constant graphon $W^{(p)} = p$, the constant function p , and, from (1.13), $W_H^{(p)}(x, y) = \frac{\binom{|V(H)|}{2}}{|Aut(H)|} p^{|E(H)|}$. It is easy to see that $W_H^{(p)}$ has only 1 non-zero eigenvalue $\lambda_1(W_H^{(p)}) = \frac{\binom{|V(H)|}{2}}{|Aut(H)|} p^{|E(H)|}$. Therefore, by Theorem 1.3,

$$\Gamma(H, K_n) \xrightarrow{D} \frac{\sigma_{H,p}}{c^{|V(H)|-1}} \cdot (\chi_{(c-1)}^2 - (c-1)). \quad (4.16)$$

where $\sigma_{H,p} := \frac{\binom{|V(H)|}{2}}{|Aut(H)|} p^{|E(H)|}$.

As another example, consider the limiting distribution in a non-symmetric example: number of monochromatic 2-stars in a complete bipartite graph.

Example 7. Let $G_n = K_{\lceil \frac{n}{2} \rceil, \lceil \frac{n}{2} \rceil}$ and $H = K_{1,2}$. Then $|Aut(H)| = 2$, and G_n converges to the graphon $W = \mathbf{1}\{(x - \frac{1}{2})(y - \frac{1}{2}) \leq 0\}$. This implies $d_W(x) = \frac{1}{2}$ for all $x \in [0, 1]$ and

$$\begin{aligned} W_{K_{1,2}}(x, y) &= \frac{W(x, y)d_W(x) + W(x, y)d_W(y) + \int_{[0,1]} W(x, z_1)W(y, z_1)dz_1}{2} \\ &= \begin{cases} \frac{1}{2} & \text{if } (x - \frac{1}{2})(y - \frac{1}{2}) \leq 0 \\ \frac{1}{4} & \text{if } (x - \frac{1}{2})(y - \frac{1}{2}) > 0. \end{cases} \end{aligned}$$

This function has two non-zero eigenvalues $\frac{3}{8}$ and $-\frac{1}{8}$ and by Theorem 1.3

$$\Gamma(K_{1,2}, K_{\lceil \frac{n}{2} \rceil, \lceil \frac{n}{2} \rceil}) \xrightarrow{D} \frac{1}{8c^2} (3\eta_1 - \eta_2),$$

where η_1 and η_2 are independent $\chi_{(c-1)}^2 - (c-1)$ random variables.

As a final example of Theorem 1.3, consider the limiting distribution of the number of monochromatic triangles in a complete tripartite graph.

Example 8. Let $G_n = K_{\lceil \frac{n}{3} \rceil, \lceil \frac{n}{3} \rceil, \lceil \frac{n}{3} \rceil}$ and $H = K_3$. Then $|Aut(H)| = 3$, and G_n converges to the graphon $W = \mathbf{1}\{(x, y) \in \mathcal{S}^c\}$, where $\mathcal{S} := [0, \frac{1}{3}]^2 \cup [\frac{1}{3}, \frac{2}{3}]^2 \cup [\frac{2}{3}, 1]^2$. A direct computation gives that for all $(u, v) \in V(K_3)$, with $u \neq v$,

$$t_{u,v}(x, y, K_3, W) = W(x, y) \int_0^1 W(x, z)W(y, z)dz = \frac{1}{3} \mathbf{1}\{(x, y) \in \mathcal{S}^c\},$$

which implies $W_{K_3}(x, y) = \frac{1}{3} \mathbf{1}\{(x, y) \in \mathcal{S}^c\}$. Now, since W_{K_3} has eigenvalues $\frac{2}{9}, -\frac{1}{9}, -\frac{1}{9}$, [Theorem 1.3](#) gives

$$\Gamma(K_3, K_{\lceil \frac{n}{3} \rceil, \lceil \frac{n}{3} \rceil, \lceil \frac{n}{3} \rceil}) \xrightarrow{D} \frac{1}{9c^2} (2\eta_1 - \eta_2 - \eta_3),$$

where η_1, η_2, η_3 are independent $\chi_{(c-1)}^2 - (c-1)$ random variables.

We conclude with an example, which shows, as before, that the condition $t(H, W) > 0$ is necessary for $\Gamma(H, G_n)$ to have a non-degenerate limit as an infinite sum of chi-squared random variables.

Example 9. Let $G_n = K_{1,n,n}$ be the complete 3-partite graph, with partitions $\{z\}, B, C$, and $H = K_3$. Given the color of the vertex z is a , using the same notations as in [Example 3](#), both $L_n(a)$ and $R_n(a)$ are independent $\text{Bin}(n, 1/c)$, and consequently

$$\sqrt{n}\Gamma(H, G_n) = \frac{T(K_3, G_n) - \mathbb{E}T(K_3, G_n)}{n^{\frac{3}{2}}} = \frac{L_n(a)R_n(a) - \frac{n^2}{c^2}}{n^{\frac{3}{2}}} \xrightarrow{D} N\left(0, \frac{2}{c} \left(1 - \frac{1}{c}\right)\right).$$

Therefore, unconditionally $\sqrt{n}\Gamma(H, G_n)$ converges to a Gaussian as well, which cannot be expressed as an infinite sum of chi-squared random variables.

Remark 4.1. A similar thing happens in [Example 4](#), where G_n is the disjoint union of the n -pyramid \mathcal{D}_n and the complete bipartite graph $K_{n,n}$ and $H = K_3$ is the triangle. In this case, it is easy to see that $T(K_3, G_n) = \frac{1}{c} \text{Bin}(n, \frac{1}{c}) + (1 - \frac{1}{c})\delta_0$, a mixture of a $\text{Bin}(n, \frac{1}{c})$ and a point mass at zero. This implies, $\Gamma(H, G_n)$ does not have a non-degenerate limiting distribution.

Acknowledgment

The authors thank the anonymous referees for providing many careful comments, which greatly improved the quality and presentation of the paper.

References

- [1] N.I. Akhiezer, *The Classical Problem of Moments and Some Related Questions of Analysis*, Oliver & Boyd, Edinburgh, 1965.
- [2] R. Arratia, L. Goldstein, L. Gordon, Poisson approximation and the Chen–Stein method, *Statist. Sci.* 5 (4) (1990) 403–424.
- [3] A.D. Barbour, L. Holst, S. Janson, *Poisson Approximations*, Oxford University Press, Oxford, 1992.
- [4] A. Basak, S. Mukherjee, Universality of the mean-field for the Potts model, *Probab. Theory Related Fields* 168 (3–4) (2017) 557–600.
- [5] T. Batu, L. Fortnow, R. Rubinfeld, W.D. Smith, P. White, Testing closeness of discrete distributions, *J. ACM* 60 (1) (2013) Article 4.
- [6] B.B. Bhattacharya, Collision times in multicolor urn models and sequential graph coloring with applications to discrete logarithms, *Ann. Appl. Probab.* 26 (6) (2016) 3286–3318.
- [7] B.B. Bhattacharya, P. Diaconis, S. Mukherjee, Universal poisson and normal limit theorems in graph coloring problems with connections to extremal combinatorics, *Ann. Appl. Probab.* 27 (1) (2017) 337–394.
- [8] B.B. Bhattacharya, S. Mukherjee, Inference in Ising models, *Bernoulli* 24 (1) (2018) 493–525.
- [9] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, K. Vesztegombi, Convergent sequences of dense graphs I: subgraph frequencies, metric properties and testing, *Adv. Math.* 219 (2009) 1801–1851.
- [10] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, K. Vesztegombi, Convergent sequences of dense graphs II. multiway cuts and statistical physics, *Ann. of Math.* 176 (2012) 151–219.
- [11] A. Cerquetti, S. Fortini, A poisson approximation for coloured graphs under exchangeability, *Sankhyā* 68 (2) (2006) 183–197.
- [12] S. Chatterjee, Stein's method and applications, UC Berkeley STAT C206A Lecture Notes, 2007, <https://statweb.stanford.edu/~souravc/Lecture6pdf>.
- [13] S. Chatterjee, P. Diaconis, Estimating and understanding exponential random graph models, *Ann. Statist.* 41 (5) (2013) 2428–2461.
- [14] S. Chatterjee, P. Diaconis, E. Meckes, Exchangeable pairs and Poisson approximation, *Electron. Encyclopedia Probab.* (2004).

- [15] S. Chatterjee, S.R.S. Varadhan, The large deviation principle for the Erdős-Rényi random graph, *European J. Combin.* 32 (7) (2011) 1000–1017.
- [16] A. DasGupta, The matching birthday and the strong birthday problem: a contemporary review, *J. Statist. Plann. Inference* 130 (2005) 377–389.
- [17] P. Diaconis, F. Mosteller, Methods for studying coincidences, *J. Amer. Statist. Assoc.* 84 (408) (1989) 853–861.
- [18] F.M. Dong, K.M. Koh, K.L. Teo, *Chromatic Polynomials and Chromaticity of Graphs*, World Scientific, Publishing Company, 2005.
- [19] X. Fang, A universal error bound in the CLT for counting monochromatic edges in uniformly colored graphs, *Electron. Commun. Probab.* 20 (2015) 1–6, Article 21.
- [20] J.H. Friedman, L.C. Rafsky, Multivariate generalizations of the Wolfowitz and Smirnov two-sample tests, *Ann. Statist.* 7 (1979) 697–717.
- [21] S.D. Galbraith, M. Holmes, A non-uniform birthday problem with applications to discrete logarithms, *Discrete Appl. Math.* 160 (10–11) (2012) 1547–1560.
- [22] T.R. Jensen, B. Toft, *Graph Coloring Problems*, in: Wiley-Interscience Series in Discrete Mathematics and Optimization, 1995.
- [23] T.R. Jensen, B. Toft, Unsolved graph coloring problems, in: L. Beineke, R. Wilson (Eds.), *Topics in Chromatic Graph Theory*, in: *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, 2015, pp. 327–357.
- [24] L. Lovász, *Large Networks and Graph Limits*, in: *Colloquium Publications*, vol. 60, 2012.
- [25] M. Nandi, D.R. Stinson, Multicollision attacks on some generalized sequential hash functions, *IEEE Trans. Inform. Theory* 53 (2) (2007) 759–767.

