MONOCHROMATIC SUBGRAPHS IN RANDOMLY COLORED GRAPHONS

BHASWAR B. BHATTACHARYA AND SUMIT MUKHERJEE

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.

1. Introduction

Let $G_n$ be a simple labeled undirected graph with vertex set $V(G_n) := \{1, 2, \ldots, |V(G_n)|\}$, 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:

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

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 $\mathbb{P}(T(K_2, G_n) > 0) = 1 - \mathbb{P}(T(K_2, G_n) = 0) = 1 - \chi_{G_n}(c_n)/c_n^{\left|V(G_n)\right|}$, where $\chi_{G_n}(c_n)$ counts the number of proper colorings of $G_n$ using $c_n$ colors. The function $\chi_{G_n}$ is known as the chromatic polynomial of $G_n$, and is a central object in graph theory [14, 18, 19]. Moreover, the statistic (1.2) shows up in various applications, for example, in the study of coincidences [13] as a generalization of the birthday paradox [2, 12], the Hamiltonian of the Ising/Potts models [3, 5], and in non-parametric two-sample tests [16]. 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. [6], for any sequence of growing graphs $G_n$.

2010 Mathematics Subject Classification. 05C15, 60C05, 60F05, 05D99.

Key words and phrases. Birthday problem, Combinatorial probability, Graph coloring, Graph limit theory, Limit theorems.

*Research partially supported by NSF grant DMS-1712037.
In this paper we consider the problem of determining the limiting distribution of the number 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

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

where:

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

$$1\{X = s\} := 1\{X_{s_1} = \cdots = X_{s_{|V(H)|}}\}, \quad (1.3)$$

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

Unlike the case for $T(K_2, G_n)$, the class of possible limiting distributions for general monochromatic subgraphs $H$ is extremely diverse [6], and, therefore, obtaining the limiting distribution of $T(H, G_n)$ for any graph sequence $G_n$ appears to be quite challenging. In this paper, we take the first step towards this goal, by providing 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 convergent sequence of dense graphs [20]. Depending on behavior of $ET(H, G_n)$ there are 3 different regimes:

1. $\mathbb{E}(T(H, G_n)) = O(1)$: In this case, $T(H, G_n)$ converges to a 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 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.

1.1. Graph Limit Theory. The theory of graph limits was developed by Lovász and coauthors [7, 8, 20], and has received phenomenal attention over the last few years. For a detailed exposition of the theory of graph limits refer to Lovász [20]. 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

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

\footnote{For a set $S$, the set $S^N$ denotes the $N$-fold cartesian product $S \times S \times \cdots \times S$.}
which is the proportion of injective maps which are homomorphisms. Moreover, denote by \( t_{\text{ind}}(F, G) \) the \textit{induced homomorphism density}, that is, the proportion of injective maps \( \phi : |V(F)| \to |V(G)| \), which satisfy \( (\phi(x), \phi(y)) \in E(G) \) if and only if \( (x, y) \in E(F) \):
\[
t_{\text{ind}}(F, G) = \frac{\sum_{s \in (V(G))(V(F))} \prod_{(a, b) \in E(F)} a_{s a s b}(G) \prod_{(a, b) \notin E(F)} (1 - a_{s a s b}(G))}{|V(G)|(|V(G)| - 1) \cdots |V(G)| - |V(F)| + 1},
\]
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 \( V(F) = \{1, 2, \ldots, |V(F)|\} \), 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.** [7, 8, 20] A sequence of graphs \( \{G_n\}_{n \geq 1} \) is said to converge to \( W \) (to be denoted by \( G_n \Rightarrow W \)) if for every finite simple graph \( F \),
\[
\lim_{n \to \infty} t(F, G_n) = t(F, W).
\]

If \( G_n \) converges to \( W \), according to definition above, then the injective homomorphism densities converge: \( t_{\text{ind}}(F, G_n) \to t(F, W) \), for every simple graph \( F \). Moreover, the induced homomorphism densities also converge, that is, \( t_{\text{ind}}(F, G_n) \to t_{\text{ind}}(F, W) \), for every simple graph \( F \), where
\[
t_{\text{ind}}(F, W) = \int_{[0,1]^{|V(F)|}} \prod_{(a, b) \in E(F)} W(x_a, x_b) \prod_{(a, b) \notin E(F)} (1 - W(x_a, x_b))dx_1 dx_2 \cdots dx_{|V(F)|}.
\]

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 \( f^G(x, y) = 1\{(|V(G)|x, |V(G)|y) \in E(G)\} \), 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 [20, 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
\[
N(H, G_n) = \frac{\text{hom}_{\text{inn}}(H, G_n)}{|\text{Aut}(H)|} \quad \text{and} \quad \mathbb{E}(T(H, G_n)) = \frac{N(H, G_n)}{c_n^{|V(H)|-1}}.
\]

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 $c_n \to \infty$, such that $\mathbb{E}T(H,G_n) \to \lambda$. Then

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

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

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). Poisson limit theorems for the number of general monochromatic subgraphs in a random coloring of a graph sequence are well-known [9, 11]. However, these results show that under general (exchangeable) coloring distribution the number of copies of any particular monochromatic subgraph converges in distribution to a Poisson. 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.

Remark 1.1. An 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 [13, Problem 3], hash-function attacks in cryptology [21], and the discrete logarithm problem [4, 17]. Refer to Example 2 for more on the birthday paradox.

1.2.2. Asymptotic Normality for Growing Colors. Theorem 1.1 asserts that if $\mathbb{E}T(H,G_n) \to \lambda$, 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 scaling by the standard deviation. Therefore, it is natural to wonder whether the same is true for $T(H,G_n)$, whenever $\mathbb{E}T(H,G_n) \to \infty$. To this end, define

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

The theorem shows that $Z(H,G_n)$ has a universal CLT whenever $\mathbb{E}T(H,G_n) \to \infty$ and $c_n \to \infty$.

Theorem 1.2. Let $G_n$ be a sequence of graphs converging to the graphon $W$, such that $t(H,W) > 0$. If $c_n \to \infty$, then\(^2\)

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

This implies $Z(H,G_n) \overset{D}{\to} N(0,1)$, whenever $\mathbb{E}T(H,G_n) \to \infty$ and $c_n \to \infty$.

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

\(^2\)The Wasserstein distance between two probability measures on $\mathbb{R}$ is, $\text{Wass}(\mu,\nu) := \sup \{ \int f \, d\nu - \int f \, d\mu : f : \mathbb{R} \to \mathbb{R} \text{ is } 1\text{-Lipschitz} \}$, that is, supremum over all $f$ such that $|f(x) - f(y)| \leq |x - y|$. 
Theorem 1.3. For the case of monochromatic edges, [6, 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 $G_n$. Error rates for the above CLT were obtained by Fang [15]. 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:

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

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

For example, when $H = K_{1,2}$ is the 2-star, with the central vertex labeled 1, then

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

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

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

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 $\{\lambda_i(K)\}_{i \geq 1}$, where every non-zero eigenvalue has finite multiplicity.

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

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

where

---

3For a graph $F = (V(F), E(F))$ and $S \subseteq V(F)$, the neighborhood of $S$ in $F$ is $N_F(S) = \{v \in V(F) : \exists u \in S$ 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.
variables, and if the coordinates of the indicators $1$ are in $\{\{X_v = a\}_{a \in [c]} : v \in V(G_n)\}$, and shows that only the quadratic term is dominant (refer to Lemmas 4.1 and 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.

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 $1\{X_v = a\}$ are replaced by independent Bernoulli variables, for every subset of vertices in $G_n$ of size $|V(H)|$. To this end, define

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

(2.1)

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

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

$$\lim_{n \to \infty} |E(T(H, G_n)^r) - E(J(H, G_n)^r)| = 0.$$

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

Proof. Let $A$ the collection of all ordered $r$-tuples $s_1, s_2, \ldots, s_r$, where $s_j \in V(G_n)^{|V(H)|}$ for $j \in [r]$, with $s_1 = (s_{11}, s_{12}, \ldots, s_{1|V(H)|})$, $s_2 = (s_{21}, s_{22}, \ldots, s_{2|V(H)|}), \ldots, s_r = (s_{r1}, s_{r2}, \ldots, s_{r|V(H)|})$, such
that \( \prod_{(a,b) \in E(H)} a_{s_a s_b} (G_n) = 1 \), for every \( j \in [r] \). Then by the multinomial expansion,

\[
\left| \mathbb{E}T(H, G_n)^r - \mathbb{E}J(H, G_n)^r \right| \leq \frac{1}{|\text{Aut}(H)|^r} \sum_A \left| \mathbb{E} \prod_{t=1}^r \{X_{s_t} \} - \mathbb{E} \prod_{t=1}^r J_{s_t} \right|,
\]

where \( F = F(s_1, \ldots, s_r) \) is the graph on vertex set \( V(F) = \bigcup_{t=1}^r s_t \) and edge set \( \bigcup_{t=1}^r \{(s_{ta}, s_{tb}) : (a, b) \in E(H)\} \), and \( b \) is the number of distinct \( |V(H)| \)-element subsets in the collection \( \{s_1, s_2, \ldots, s_r\} \). Note that if the graph \( F \) is connected, \( |V(F)| - 1 \leq b|V(H)| - b \), and therefore, in general \( |V(F)| - \nu(F) \leq b|V(H)| - b \), where \( \nu(F) \) is the number of connected components of \( F \).

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

\[
\left| \mathbb{E}T(H, G_n)^r - \mathbb{E}J(H, G_n)^r \right| \lesssim_H \sum_{(p,q): q|V(H)| < p \leq r|V(H)|} \sum_{\mathcal{N}_{p,q,r}} \frac{1}{c_n^{p-q}}
\]

\[
\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{1}{|V(G_n)|^T(p|V(H)|^{-1})^{p|V(H)|^{-1}}} \quad \text{(since \( c_n^{V(H)|^{-1}} \approx \Theta(|V(G_n)|^{|V(H)|}))}
\]

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 \( n \), (2.2) follows.

Finally, from the above arguments it also follows that

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

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

\[
J(H, G_n) \to \sum_{F \supseteq H, |V(F)| = |V(H)|} N(H, F) X_F,
\]

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

\[
J(H, G_n) = \frac{1}{|Aut(H)|} \sum_{s \in V(G_n) \setminus V(H)} \prod_{(a,b) \in E(H)} a_{s_a s_b}(G_n) J_s
= \sum_{s \in \binom{V(G_n)}{|V(H)|}} N(H, G_n[s]) J_s
= \sum_{F \supseteq H : |V(F)| = |V(H)|} N(H, F) \sum_{s \in \binom{V(G_n)}{|V(H)|}} 1\{G_n[s] = F\} \cdot J_s \tag{2.3}
\]

Now, note that, by definition, the collection

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

is independent and for every fixed \( F \), \( J_n(F) := \sum_{s \in \binom{V(G_n)}{|V(H)|}} 1\{G_n[s] = F\} J_s \) is a sum of independent Bernoulli \( \left(1, \frac{1}{c_n^{V(H)} - 1}\right) \) random variables. Therefore, to prove theorem it suffices to show that \( J_n(F) \to X_F \) (with \( X_F \) as defined in the statement of the theorem) in distribution and in moments, which follows if we can prove that

\[
\mathbb{E}(J_n(F)) \to \lambda \cdot \frac{|Aut(H)|}{|Aut(F)|} \cdot \frac{\tind(F, W)}{t(H, W)}. \tag{2.4}
\]

To show (2.4), first note that

\[
\frac{|V(G_n)| |V(H)|}{c_n^{V(H)} - 1} = N(H, G_n) \cdot \frac{|V(G_n)| |V(H)|}{c_n^{V(H)} - 1} = (1 + o(1)) \lambda \cdot \frac{|V(G_n)| |V(H)|}{\text{hominj}(H, G_n) / |Aut(H)|} \to \lambda \frac{|Aut(H)|}{t(H, W)}.
\]

Then recalling (1.7),

\[
\mathbb{E} \sum_{s \in \binom{V(G_n)}{|V(H)|}} 1\{G_n[s] = F\} \cdot J_s
= \frac{1}{c_n^{V(H)} - 1} \sum_{s \in \binom{V(G_n)}{|V(H)|} \setminus (a,b) \in E(F)} \prod_{(a,b) \in E(F)} a_{s_a s_b}(G_n) \prod_{(a,b) \notin E(F)} (1 - a_{s_a s_b}(G_n))
= \frac{|V(G_n)| |V(H)|}{c_n^{V(H)} - 1} \sum_{s \in \binom{V(G_n)}{|V(H)|}} \prod_{(a,b) \notin E(F)} a_{s_a s_b}(G_n) \prod_{(a,b) \in E(F)} (1 - a_{s_a s_b}(G_n))
= (1 + o(1)) \lambda \frac{|Aut(H)|}{t(H, W)} \cdot \frac{1}{|V(G_n)| |V(H)|} \sum_{s \in \binom{V(G_n)}{|V(H)|}} \prod_{(a,b) \notin E(F)} a_{s_a s_b}(G_n) \prod_{(a,b) \in E(F)} (1 - a_{s_a s_b}(G_n))
= (1 + o(1)) \lambda \frac{|Aut(H)|}{t(H, W)} \cdot \frac{\tind(F, G_n)}{|Aut(F)|} \tag{recall (1.4)}
\]
\[ \lambda \frac{|\text{Aut}(H)|}{t(H,W)} \cdot t_{\text{ind}}(F,W) \cdot \frac{\text{ind}(F,W)}{|\text{Aut}(F)|}, \]  
where the last step uses \( t_{\text{ind}}(F,G_n) \to t_{\text{ind}}(F,W) \) (since \( G_n \) converges to \( W \)). \hfill \Box

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 Lemma 2.1 and Lemma 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 \( T(H,G_n) \overset{D}{\to} Y \) 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 \( t(H,W(p)) = p^{\text{ind}(E(H))} \) and \( t_{\text{ind}}(F,W(p)) = p^{\text{ind}(E(F))} (1 - p)^{|E(F)|} \). Therefore, by Theorem 1.1, choosing \( c_n \) such that \( \mathbb{E}T(H,G_n) = (1 + o_p(1)) \frac{\lambda}{c_n|V(H)|^{s-1}} \to \lambda \), gives

\[ T(H,G_n) \overset{D}{\to} \sum_{F \supseteq H: |V(F)| = |V(H)|} N(H,F)X_F, \]

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

- When \( G_n = K_n \) (that is, \( p = 1 \)), \( X_{K_n} \sim \text{Pois} \left( \lambda \cdot \frac{|\text{Aut}(H)|}{|V(H)|} \right) = \text{Pois} \left( \frac{\lambda}{N(H,K_n)} \right) \)

\[ T(H,K_n) \overset{D}{\to} N(H,K_n) \cdot \text{Pois} \left( \frac{\lambda}{N(H,K_n)} \right). \]

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

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 \( T(K_s,G_n) \overset{D}{\to} \text{Pois}(\lambda) \), where \( \frac{N(K_s,G_n)}{c_n^{s-1}} \to \lambda \), and \( \frac{1}{|V(G_n)|} N(K_s,G_n) \to \frac{1}{s!} t(K_s,W) \), gives

\[ \mathbb{P}(T(K_s,G_n) > 0) \approx 1 - \exp \left( - \frac{N(K_s,G_n)}{c_n^{s-1}} \right) = p \implies |V(G_n)| \approx \left( \frac{s!}{t(K_s,W)} c_n^{s-1} \log \left( \frac{1}{1 - p} \right) \right)^{\frac{1}{s}} , \]

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
Given a 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 \( c_n^{\left| V(H) \right| - 1} = \Theta(\left| V(G_n) \right|^{\left| V(H) \right|}) \), 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 \( c_n^{\left| V(H) \right| - 1} \sim N(H, G_n) \) 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 \( W(x, y) = 1\{(x - \frac{1}{2})(y - \frac{1}{2}) \leq 0\} \), for which \( t(K_3, W) = 0 \). However, if we color \( G_n \) randomly with \( c_n = n \) colors, such that \( N(K_3, G_n)/c_n^3 = 1 \), 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,

\[
L_n(a) \sim \text{Bin}(n, 1/c_n) \overset{D}{\to} X, \quad R_n(a) \sim \text{Bin}(n, 1/c_n) \overset{D}{\to} Y,
\]

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

### 3. Asymptotic Normality: Proof of Theorem 1.2

In this section we prove the asymptotic normality of \( Z(H, G_n) \) (recall (1.9)), whenever \( \mathbb{E}T(H, G_n) \to \infty \) such that \( c_n \to \infty \). We begin the following definition.

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

\[
E(H^2_{(a, b)}) = E(H) \cup E(H' \setminus \{a', b'\}) \bigcup \{(a, a') : x' \in N_{H'}(a')\} \bigcup \{(b, y') : y' \in N_{H'}(b')\}.
\]

Note that \( H^2_{(a, b)} \) has \( 2|V(H)| - 2 \) edges 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 \( \{1, 2, \ldots, |V(H)|\} \) such that \( a, b \in V(H) \). Then \( t(H^2_{(a, b)}, W) > 0 \), whenever \( t(H, W) > 0 \).

**Proof.** Recalling (1.2), we have

\[
t(H, W) = \int_{[0, 1]^2} t_{a, b}(x_a, x_b, H, W)dx_a dx_b
\]

\[
\leq \left( \int_{[0, 1]^2} t^2_{a, b}(x_a, x_b, H, W)dx_a dx_b \right)^{\frac{1}{2}} \leq \left( t(H^2_{(a, b)}, W) \right)^{\frac{1}{2}} \quad \text{(by Cauchy-Schwarz)}
\]

This implies \( t(H^2_{(a, b)}, W) \geq t(H, W) > 0 \). \( \square \)
Definition 3.2. For \( s \in V(G_n)_{|V(H)|} \), define
\[
Z_s := 1\{X_{=s}\} - \mathbb{E}1\{X_{=s}\} = 1\{X_{=s}\} - \frac{1}{c_n^{|V(H)|-1}},
\]
where \( 1\{X_{=s}\} \) is as defined in (1.3). Then
\[
T(H, G_n) - \mathbb{E}T(H, G_n) = \sum_{s \in V(G_n)_{|V(H)|}} M_{G_n}(s, H) \left( 1\{X_{=s}\} - \frac{1}{c_n^{|V(H)|-1}} \right)
\]
\[
= \sum_{s \in V(G_n)_{|V(H)|}} M_{G_n}(s, H)Z_s,
\]
where \( M_{G_n}(s, H) = \frac{1}{|\text{Aut}(H)|} \prod_{(a, b) \in E(H)} a_{s(a)s(b)}(G_n) \).

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. For \( s \in V(G_n)_{|V(H)|} \), let \( 1\{X_{=s}\} \) and \( Z_s \) be as defined above. Then the following hold:

(a) For \( s, t \in V(G_n)_{|V(H)|} \),
\[
\mathbb{E}1\{X_{=s}\}1\{X_{=t}\} = \begin{cases} 
\frac{1}{c_n^{|s \cup t|-1}} & \text{if } |s \cup t| = 2|V(H)| \\
\frac{1}{c_n^{|s \cup t|-1}} & \text{if } |s \cup t| \leq 2|V(H)| - 1.
\end{cases}
\]

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

(c) If \( t(H, W) > 0 \), then
\[
\text{Var}(T(H, G_n)) \geq \max \left( \frac{|V(G_n)|^{|V(H)|}}{c_n^{|V(H)|-1}}, \frac{|V(G_n)|^{2|V(H)|-2}}{c_n^{2|V(H)|-3}} \right).
\]

Proof. If \( |s \cup t| = 2|V(H)| \), the indices \( s, t \) do not intersect. In this case, the expectation factorizes by independence, and
\[
\mathbb{E}1\{X_{=s}\}1\{X_{=t}\} = \mathbb{E}1\{X_{=s}\}\mathbb{E}1\{X_{=t}\} = \frac{1}{c_n^{2|V(H)|-2}}.
\]

Otherwise,
\[
\mathbb{E}1\{X_{=s}\}1\{X_{=t}\} = \mathbb{P}(X_{s_1} = \cdots, X_{s_{|V(H)|}} = X_{t_1} = \cdots = X_{t_{|V(H)|}}) = \frac{1}{c_n^{|s \cup t|-1}},
\]
completing the proof of (a). The result in (b) follows from (a) and observing that \( \mathbb{E}1\{X_{=s}\}\mathbb{E}1\{X_{=t}\} = \frac{1}{c_n^{2|V(H)|-2}} \).

To show (c) note that
\[
\text{Var}(T(H, G_n)) = \sum_{s \in V(G_n)_{|V(H)|}} M_{G_n}(s, H)\mathbb{E}Z_s^2 + \sum_{s \neq t \in V(G_n)_{|V(H)|}} M_{G_n}(s, H)M_{G_n}(t, H)\mathbb{E}Z_s Z_t.
\]
Now, since each of the terms in the covariance is non-negative by part (b),

\[ \text{Var}(T(H, G_n)) \geq \sum_{s \in V(G_n) \setminus V(H)} M_{G_n}(s, H) \mathbb{E}Z^2_s \geq \frac{N(H, G_n)}{c_n^{2|V(H)|-3}} \geq \frac{|V(G_n)|^{2|V(H)|-2}}{c_n^{2|V(H)|-3}}, \]

since \( \frac{N(H, G_n)}{|V(G_n)|^{2|V(H)|-3}} = (1 + o(1)) \frac{1}{\text{Aut}(H)} \to t(H, W) > 0 \), when \( G_n \) converges to \( W \). Next, considering the sum over pairs \( s, t \in V(G_n) \setminus V(H) \) such that \( s = (s_1, s_2, s_3, \ldots, s_{|V(H)|}), t = (t_1, t_2, t_3, \ldots, t_{|V(H)|}) \) with the indices \( \{s_1, s_2, s_3, \ldots, s_{|V(H)|}, t_1, t_2, t_3, \ldots, t_{|V(H)|}\} \) all distinct

\[ \text{Var}(T(H, G_n)) \geq \frac{N(H^2, G_n)}{c_n^{2|V(H)|-3}} \geq \frac{|V(G_n)|^{2|V(H)|-2}}{c_n^{2|V(H)|-3}}, \]

where the last step uses \( \lim_{n \to \infty} \frac{1}{|V(G_n)|^{2|V(H)|-3}} N(H^2, G_n) \gtrsim t(H^2, W) > 0 \) by Lemma 3.1. Combining these two estimates give the desired lower bound on the variance.

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

**Lemma 3.3.** Let \( s_1, s_2, s_3 \in V(G_n) \setminus V(H) \) 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 \( \mathbb{E}[Z_{s_1}Z_{s_2}Z_{s_3}] \leq \frac{8}{c_n^{3|V(H)|-1}} \).

**Proof.** To begin with we show that \( |s_2 \cap (s_1 \cup s_3)| \leq 1 \). To see this, observe

\[ 3|V(H)| - 2 = \sum_{a=1}^{3} |s_a| = |V(H)| + |s_1 \cup s_3| - (s_1 \cup s_3) \cap s_2 \leq |V(H)| + 2|V(H)| - 1 - (s_1 \cup s_3) \cap s_2, \]

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

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

since \( \mathbb{P}(X_j = X_{q_2} \cdots = X_{q_{|V(H)|}} | X_j) = \frac{1}{c_n^{3|V(H)|-1}} \), for any \( q_2, q_3, \ldots, q_{|V(H)|} \in V(G_n) \) distinct. This completes the proof of (a).

Next, by a direct expansion, it follows that

\[ \mathbb{E}[Z_{s_1}Z_{s_2}Z_{s_3}] \leq \mathbb{E}[1\{X_{s_1}\}1\{X_{s_2}\}1\{X_{s_3}\} + \frac{4}{c_n^{3|V(H)|-3}} + \frac{1}{c_n^{3|V(H)|-3}} \sum_{1 \leq a < b \leq 3} \mathbb{E}[1\{X_{s_a}\}1\{X_{s_b}\} \mathbb{E}[1\{X_{s_c}\}], \]  

(3.3)

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

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

(3.4)
Thus, combining (3.5) and (3.6), \( \text{Cov}(Z, s) \) holds for the pairs \((s_1, s_3)\) as well, and, therefore, the RHS of (3.3) is bounded by \( \frac{8}{c_n} |s_1 \cup s_3| - 1 \).

**Lemma 3.4.** Let \( s_1, s_2, s_3, s_4 \in V(G_n) \) 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 \( \text{Cov}(Z_{s_1}, Z_{s_2}, Z_{s_3}, Z_{s_4}) \leq \frac{16}{c_n} |s_1 \cup s_3| - 1 \).

**Proof.** If \( |\bigcup_{a=1}^4 s_a| = 4|V(H)| - 2 \), then the index sets \( s_1 \cup s_2 \) and \( s_3 \cup s_4 \) are distinct, 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 the index sets are not disjoint, then we must have \( |s_1 \cup s_2| = |s_3 \cup s_4| = 2| V(H) | - 1 \), which implies

\[
\mathbb{E}Z_{s_1} Z_{s_2} Z_{s_3} Z_{s_4} = 0 \tag{3.5}
\]

by Lemma 3.2(a). Moreover, in this case \( |(s_1 \cup s_2) \cap (s_3 \cup s_4)| = 1 \), which implies that one of the sets \( s_1 \cap (s_3 \cup s_4) \) or \( s_2 \cap (s_3 \cup s_4) \) is empty. Assuming, without loss of generality, that \( s_1 \cap (s_3 \cup s_4) \) is empty, implies \( s_2 \cap (s_1 \cup s_3 \cup s_4) = s_2 \cap s_1 = \{j\} \) is a singleton. Then

\[
\mathbb{E}Z_{s_1} Z_{s_2} Z_{s_3} Z_{s_4} = \mathbb{E}\{ \mathbb{E}(Z_{s_2}) X_j, \{X_i, i \in \{V(G_n) \} \}} Z_{s_1} Z_{s_2} Z_{s_4} = 0. \tag{3.6}
\]

Combining (3.5) and (3.6), \( \text{Cov}(Z_{s_1}, Z_{s_2}, Z_{s_3}, Z_{s_4}) = 0 \), whenever \( |\bigcup_{a=1}^4 s_a| = 4|V(H)| - 3 \). This completes the proof of (a).

To show (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 bound up to fourth joint moments of \( \mathbb{1}\{ X = s \} \). To this end, note that \( |\bigcup_{a=1}^4 s_a| \cap |s_3 \cap s_4| \geq 1 \) which gives

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

Thus,

\[
\frac{1}{c_n^{-1}} \mathbb{E}\left\{ \left( \bigcup_{a=1}^4 s_a \right) \right\} 1 \{X = s_1\} 1 \{X = s_2\} 1 \{X = s_3\} 1 \{X = s_4\} = \frac{1}{c_n^{-1}} \mathbb{E}\left\{ \left| \bigcup_{a=1}^3 s_a \right| + |V(H)| - 2 \right\} \leq \frac{1}{c_n^{-1}} \mathbb{E}\left\{ \left| \bigcup_{a=1}^4 s_a \right| - |V(H)| - 2 \right\}, \tag{3.7}
\]

and a similar bound holds for all other three triples.

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
\[
\frac{1}{c_n^{2|V(H)|-2}} \mathbb{E} \{ X_{s_1} \} \mathbb{E} \{ X_{s_2} \} \leq \frac{1}{c_n^{2|V(H)|-2}} \max \left( \frac{1}{c_n^{2|V(H)|-2}}, \frac{1}{|s_1 \cup s_2|} \right) \leq \max \left( \frac{1}{c_n^{4|V(H)|-4}}, \frac{1}{c_n^{2|U_{a=1}^n s_a|-1}} \right),
\]
and a similar bound applies for all the other two tuples. Thus, expanding the fourth moment and using (3.7) and (3.8) gives
\[
\mathbb{E} |X_{s_1} X_{s_2} X_{s_3} X_{s_4}| \leq \frac{9}{c_n^{2|U_{a=1}^n s_a|-1}} + \frac{6}{c_n^{4|V(H)|-4}} \leq \frac{15}{c_n^{2|U_{a=1}^n s_a|-1}}.
\]
Finally, by Lemma 3.2(b), if \( \max(|s_1 \cup s_2|, |s_3 \cup s_4|) \geq 2|V(H)| - 1 \), then \( \mathbb{E} X_{s_1} X_{s_2} \mathbb{E} X_{s_3} X_{s_4} = 0 \). 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
\[
|\mathbb{E} X_{s_1} X_{s_2} \mathbb{E} X_{s_3} X_{s_4}| \leq \frac{1}{c_n^{2|s_1 \cup s_2|+s_3 \cup s_4|-2}} \leq \frac{1}{c_n^{2|U_{a=1}^n s_a|-1}},
\]
where the last inequality uses the fact that \( |(s_1 \cup s_2) \cap (s_3 \cup s_4)| \geq 1 \). Combining (3.9) along with (3.10) completes the proof of the lemma. \( \Box \)

**Proof of Theorem 1.2.** Recall \( Z(H, G_n) \) from (1.9). The CLT for \( Z(H, G_n) \) will be proved using the Stein’s method based on dependency graphs. To this end, for every \( s_1 \in V(G_n)|V(H)| \) let
\[
N_{s_1} := \{ s_2 \in V(G_n)|V(H)| : |s_1 \cap s_2| \geq 1 \}.
\]
In words \( N_{s_1} \) is the subset of tuples in \( V(G_n)|V(H)| \) which have at least one index common with \( s_1 \). Then we have (see [10])
\[
\text{Wass} (Z(H, G_n), N(0,1)) \leq R_1 + R_2,
\]
where
\[
R_1 = \left( K_0 \text{Var} \left( \frac{\sum_{s_1 \in V(G_n)|V(H)|} X_{s_1} \sum_{s_2 \in N_{s_1}} X_{s_2}}{\sigma_n^2} \right) \right)^{1/2},
\]
\[
R_2 = \frac{\sum_{s_1 \in V(G_n)|V(H)|} \mathbb{E} |X_{s_1}| \left( \sum_{s_2 \in N_{s_1}} X_{s_2} \right)^2}{\sigma_n^3},
\]
with \( \sigma_n^2 = \text{Var}(T(H, G_n)) \) and \( K_0 = \sqrt{2/\pi} \).

We will bound each of the terms above separately. To begin with, observe
\[
\text{Var} \left( \frac{1}{\sigma_n^2} \sum_{s_1 \in V(G_n)|V(H)|} X_{s_1} \sum_{s_2 \in N_{s_1}} X_{s_2} \right)
\]
\[
= \frac{1}{\sigma_n^4} \sum_{s_1 \in V(G_n)|V(H)|} \sum_{s_2 \in N_{s_1}} \sum_{s_3 \in V(G_n)|V(H)|} \sum_{s_4 \in N_{s_3}} \text{Cov}(X_{s_1} X_{s_2}, X_{s_3} X_{s_4}).
\]
Let \( \ell = |\bigcup_{a=1}^n s_a| \) 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 (up to constants depending on \( |V(H)| \)) is given by
\[
\frac{1}{\sigma_n^4} \sum_{\ell = |V(H)|}^{4|V(H)|-4} \frac{|V(G_n)|^\ell}{c_n^{\ell-1}} \leq \frac{1}{\sigma_n^4} \left( \frac{|V(G_n)|^{4|V(H)|-4}}{c_n^{4|V(H)|-5}} + \frac{|V(G_n)|^{4|V(H)|-4}}{c_n^{4|V(H)|-5}} \right).
\]
Again let \( \ell \) be the number of vertices in \( V(H) \). Therefore, \( Z \) and \( \ell \) are independent, and it follows that the first term in the RHS of Theorem 1.2 goes to zero whenever \( E(H,G) \rightarrow \infty \).

Combining (3.11) with (3.12) and (3.13) completes the proof of (1.10).

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

\[
E(D_n) = \{(a,b), (a,c_1), (a,c_2), \ldots, (a,c_n), (b,c_1), (b,c_2), \ldots, (b,c_n)\}.
\]
Lemma 4.1. For every $v$

Proof.

Observation 4.1. For $s \in V(G_n)\setminus V(H)$, let $Z_s = 1\{X_s = 1\} - \frac{1}{c}$. Then

$$Z_s = \sum_{a=1}^{c} \sum_{J \subseteq V(H)} \frac{1}{e^{V(H)} - |J|} \prod_{j \in J} \left(1\{X_j = a\} - \frac{1}{c}\right).$$

Proof. This follows by directly multiplying out the RHS above, and observing that, for every $v \in V(G_n)$, $\sum_{a=1}^{c} (1\{X_v = a\} - \frac{1}{c}) = 0$.

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

$$T(H, G_n) - \mathbb{E}(T(H, G_n)) = \sum_{s \in V(G_n) \setminus V(H)} M_{G_n}(s, H)Z_s \quad \text{(recall (3.2))}$$

$$= \sum_{s \in V(G_n) \setminus V(H)} M_{G_n}(s, H) \sum_{a=1}^{c} \sum_{J \subseteq V(H)} \frac{1}{e^{V(H)} - |J|} \prod_{j \in J} \left(1\{X_j = a\} - \frac{1}{c}\right)$$

$$= \sum_{\substack{J \subseteq V(H) \\ |J| \geq 2}} T_J(H, G_n), \quad (4.1)$$

where

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

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

$$\mathbb{E}T_J(H, G_n) = \sum_{a=1}^{c} \sum_{s \in V(G_n) \setminus V(H)} M_{G_n}(s, H) \frac{1}{e^{V(H)} - |J|} \mathbb{E} \left(\prod_{j \in J} 1\{X_j = a\} - \frac{1}{c}\right) = 0, \quad (4.2)$$

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 $s \in V(G_n)\setminus V(H)$, let $Z_s = 1\{X_s = 1\} - \frac{1}{c}$. Then

$$Z_s = \sum_{a=1}^{c} \sum_{J \subseteq V(H)} \frac{1}{e^{V(H)} - |J|} \prod_{j \in J} \left(1\{X_j = a\} - \frac{1}{c}\right).$$

Proof. This follows by directly multiplying out the RHS above, and observing that, for every $v \in V(G_n)$, $\sum_{a=1}^{c} (1\{X_v = a\} - \frac{1}{c}) = 0$.

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

$$T(H, G_n) - \mathbb{E}(T(H, G_n)) = \sum_{s \in V(G_n) \setminus V(H)} M_{G_n}(s, H)Z_s \quad \text{(recall (3.2))}$$

$$= \sum_{s \in V(G_n) \setminus V(H)} M_{G_n}(s, H) \sum_{a=1}^{c} \sum_{J \subseteq V(H)} \frac{1}{e^{V(H)} - |J|} \prod_{j \in J} \left(1\{X_j = a\} - \frac{1}{c}\right)$$

$$= \sum_{\substack{J \subseteq V(H) \\ |J| \geq 2}} T_J(H, G_n), \quad (4.1)$$

where

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

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

$$\mathbb{E}T_J(H, G_n) = \sum_{a=1}^{c} \sum_{s \in V(G_n) \setminus V(H)} M_{G_n}(s, H) \frac{1}{e^{V(H)} - |J|} \mathbb{E} \left(\prod_{j \in J} 1\{X_j = a\} - \frac{1}{c}\right) = 0, \quad (4.2)$$
since
\[
\mathbb{E} \left( \prod_{j \in J} 1\{X_{s_j} = a\} - \frac{1}{c} \right) = \prod_{j \in J} \mathbb{E} \left( 1\{X_{s_j} = a\} - \frac{1}{c} \right) = 0,
\]
since \{\{1\{X_{s_j} = a\} - \frac{1}{2} : j \in J\}\} is a collection of independent random variables.

The second moment of \(T_J(H, G_n)\) equals
\[
\sum_{a, a' \in [d]} \sum_{s, s' \in V(G_n)\backslash V^H} M_{G_n}(s, H) M_{G_n}(s', H) \frac{1}{c^{|V(H)|-2|J|}} \mathbb{E} \prod_{j \in J} \left( 1\{X_{s_j} = a\} - \frac{1}{c} \right) \left( 1\{X_{s'_j} = a'\} - \frac{1}{c} \right)
\]
(4.3)

If there exists \(s_0 \in \{s_j : j \in J\}\backslash\{s'_j : j \in J\}\), then
\[
\mathbb{E} \prod_{j \in J, s_j \neq s_0} \left( 1\{X_{s_j} = a\} - \frac{1}{c} \right) \prod_{j \in J} \left( 1\{X_{s'_j} = a'\} - \frac{1}{c} \right) \mathbb{E} \left( 1\{X_{s_0} = a\} - \frac{1}{c} \right) = 0.
\]
Similarly, the expectation vanishes if \(s_0 \in \{s'_j : j \in J\}\backslash\{s_j : j \in J\}\). Therefore, (4.3) gives
\[
\mathbb{E} T_J(H, G_n)^2 \leq \sum_{s, s' \in V(G_n)\backslash V^H, \{s_j : j \in J\} = \{s'_j : j \in J\}} M_{G_n}(s, H) M_{G_n}(s', H)
\]
(4.4)
\[
= O(\|V(G_n)\|^2 |V(H)| - |J|) = o(\|V(G_n)\|^2 |V(H)| - 2),
\]
whenever \(|J| \geq 3\).

Combining (4.2) and (4.4) it follows that \(T_J(H, G_n) = o_P(\|V(G_n)\| |V(H)|^{-1})\), whenever \(|J| \geq 3\).

**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 \(\phi : V(H) \to V(G_n)\) such that \(\phi(u) = i\) and \(\phi(v) = j\). More formally,
\[
M_{u,v}(i, j, H, G_n) = a_{ij, uv}(G_n) \sum_{s \in V(G_n)\backslash V^H} \prod_{x \in N_H(u) \setminus \{v\}} a_{ix}(G_n) \prod_{y \in N_H(v) \setminus \{u\}} a_{jy}(G_n) \prod_{(x, y) \in E(H \setminus \{u, v\})} a_{xy}(G_n),
\]
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 \(s \setminus \{s_u, s_v\}\), with \(s \in V(G_n)\backslash V^H\), which are distinct and belong to \([|V(G_n)|]|\{i, j\}\). Note that \(M_{u,v}(i, j, H, G_n)\) is, in general, not symmetric in \(i, j\), but satisfies \(M_{v,u}(i, j, H, G_n) = M_{u,v}(j, i, H, G_n)\).

Finally, define the symmetric scaled 2-point homomorphism matrix as ((\(\overline{B}_H(G_n)_{ij}\))) for \(1 \leq i \neq j \leq |V(G_n)|\) with
\[
\overline{B}_H(G_n)_{ij} := \frac{1}{2|\text{Aut}(H)| \cdot |V(G_n)| |V(H)| - 1} \sum_{1 \leq u \neq v \leq |V(H)|} M_{u,v}(i, j, H, G_n),
\]
(4.5)
for \(1 \leq i \neq j \leq |V(G_n)|\).

\(\footnote{For example, when \(H = K_{1,2}\) is the 2-star with the central vertex labeled 1, then \(M_{1,2}(i, j, K_{1,2}, G_n) = M_{1,2}(i, j, K_{1,2}, G_n) = a_{ij}(G_n) \cdot (d_G(i) - a_{ij}(G_n)), \) where \(d_G(i)\) 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 \(i, j\).}
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.** Define,

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

where $B_H(G_n)$ is the 2-point homomorphism matrix as defined in (4.5). Then, $\Gamma(H, G_n) = \Gamma_2(H, G_n) + o_P(1)$.

**Proof.** Note that

$$c|V(H)|^{-2} \sum_{J \subseteq V(H)} T_J(H, G_n)$$

$$= \sum_{1 \leq u < v \leq |V(H)|} \frac{M_{u,v}(s_u, s_v, H, G_n)}{|Aut(H)|} \left( 1\{X_u = a\} - \frac{1}{c} \right) \left( 1\{X_v = a\} - \frac{1}{c} \right),$$

$$= \sum_{1 \leq u < v \leq |V(H)|} \frac{M_{u,v}(i, j, H, G_n)}{|Aut(H)|} \left( 1\{X_i = a\} - \frac{1}{c} \right) \left( 1\{X_j = a\} - \frac{1}{c} \right),$$

$$= |V(G_n)|^{V(H)|-1} \sum_{1 \leq i \neq j \leq |V(G_n)|} B_H(G_n)_{ij} \left( 1\{X_i = a\} - \frac{1}{c} \right) \left( 1\{X_j = a\} - \frac{1}{c} \right),$$

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

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

Now, from (4.5) it is easy to see that $B_H(G_n)_{ij} = \frac{B_H(G_n)_{ij} + B_H(G_n)_{ji}}{2}$, 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 $\{R_v : v \in V(G_n)\}$, where $R_v = (1\{X_v = a\} - \frac{1}{c} a \in [c])$, are replaced by a collection of i.i.d. Gaussian vectors with the same mean and covariance structure. More formally,

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

with $\hat{U}_{v,a} = U_{v,a} - \overline{U}_v$, 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 $\overline{U}_v = \frac{1}{c} \sum_{a=1}^{c} U_{v,a}$. Note that for each $v \in V(G_n)$ the random vector $\hat{U}_v := (\hat{U}_{v,1}, \hat{U}_{v,2}, \ldots, \hat{U}_{v,c})$ has mean 0 and the same covariance matrix as $R_v$. Also, $\{\hat{U}_v, v \in V(G_n)\}$ are independent and identically distributed random vectors. Finally, define

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

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 $r \geq 1$,

$$\lim_{n \to \infty} \|E\Gamma_2(H, G_n)^r - E\Delta_2(H, G_n)^r\| = 0.$$
Figure 1 shows a 5-cycle of the image of the vertex $z$ in the graph obtained by the union of $H$. Denote this graph by $G$. From Definition 1.2, it is easy to see that the limiting distribution of $\Gamma(H, G)$ is the sum of $c$-quadratic forms in $\mathbf{B}_H(G)$. To this end, we need to understand the spectrum of the matrix $\mathbf{B}_H(G)$. We begin by defining the notion of cycles formed by $H$, which arise in the analysis of the power-sum of the eigenvalues of $\mathbf{B}_H(G)$.

**Definition 4.2.** Fix an integer $g \geq 2$, and let $H_1, H_2, \ldots, H_g$ 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 $H_1, H_2, \ldots, H_g$, 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 $v_{g+1}^{(1)} := u_1^{(1)}$ and $H_{g+1} = H_1$. Denote this graph by $H^{(g)}(J)$. From Definition 1.2, it is easy to see that

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

Figure 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 (1.12), we proceed to prove the convergence of the spectrum of $\mathbf{B}_H(G)$.
Lemma 4.4. Let \( \lambda_1(\mathcal{B}_H(G_n)) \geq \lambda_2(\mathcal{B}_H(G_n)) \geq \cdots \geq \lambda_{|V(G_n)|}(\mathcal{B}_H(G_n)) \) be the eigenvalues of \( \mathcal{B}_H(G_n) \). Then, for every \( g \geq 2 \), \( \lim_{n \to \infty} \sum_{r=1}^{V(G_n)} \lambda_r(\mathcal{B}_H(G_n))^g = \sum_{i=1}^{\infty} \lambda_r(H,W)^g \), where \( \lambda_1(H,W) \geq \lambda_2(H,W) \geq \cdots \geq \lambda_{|V(G_n)|}(H,W) \) are the eigenvalues of \( W_H \). Moreover, the assumption \( t(H,W) > 0 \) ensures that at least one eigenvalue of \( W_H \) is non-zero.

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

\[
\begin{align*}
\sum_{r=1}^{|V(G_n)|} \lambda_r(\mathcal{B}_H(G_n))^g &= \text{tr}(\mathcal{B}_H(G_n))^g) \\
&= \sum_{j \in V(G_n)g} \prod_{a=1}^g \mathcal{B}_H(G_n)_{ja_ja+1} \\
&= K_g \cdot \frac{1}{|V(G_n)|^g|V(H)|^{-g}} \sum_{j \in V(G_n)g} \prod_{a=1}^g \sum_{1 \leq u \neq v \leq |V(H)|} M_{u,v}(j_{a_ja+1}, H, G_n) \\
&= K_g \cdot \frac{1}{|V(G_n)|^g|V(H)|^{-g}} \sum_{1 \leq u_1 \neq v_1 \leq |V(H)|} \cdots \sum_{1 \leq u_g \neq v_g \leq |V(H)|} \prod_{j \in V(G_n)g} M_{u_v,u_{a_ja+1}}(j_{a_ja+1}, H, G_n).
\end{align*}
\]

Now, fix \( J = \{(u_a, v_a) : 1 \leq u_a \neq v_a \leq |V(H)|, a \in [g]\} \). Recalling Definition 4.1, the product in

\[
\sum_{j \in V(G_n)g} \prod_{a=1}^g M_{u_v,u_{a_ja+1}}(j_{a_ja+1}, H, G_n)
\]
can be expanded to obtain a sum over at most \( g|V(H)| - g \) indices in \( V(G_n) \). However, if any two of the indices are the same, then the corresponding term in the sum is \( o(|V(G_n)|^g|V(H)|^{-g}) \). Therefore, the leading term is a sum over \( g|V(H)| - g \) distinct indices ranging in \( V(G_n) \), which is 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,

$$
\frac{1}{|V(G_n)|^{2g}|V(H)|^{g}} \sum_{j \in V(G_n)} \prod_{a=1}^{g} M_{u_a,v_a}(j_a, j_{a+1}, H, G_n) = t(H^{(g)}(J), G_n) + o(1) \to t(H^{(g)}(J), W).
$$

Next, recall that $\lambda_1(H, W) \geq \lambda_2(H, W) \geq \cdots$ are the eigenvalues of the function $W_H$, as defined in Theorem 1.3. Finally, a similar calculation as above gives

$$
\sum_{r=1}^{\infty} \lambda^2_r(H, W) = K_g \sum_{1 \leq u_1 \neq v_1 \leq |V(H)|} \cdots \sum_{1 \leq u_g \neq v_g \leq |V(H)|} t(H^{(g)}(J), W),
$$

and so the $g$-th power sum of the eigenvalues of $\mathcal{B}_H(G_n)$ 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 $1 \leq a \neq b \leq |V(H)|$, $H^{(2)}(J) = H^{(2)}_{(a,b)}$ (recall Definition 3.1) and by Lemma 3.1 $t(H^{(g)}(J), W) > 0$. Therefore, $\sum_{r=1}^{\infty} \lambda^2_r(H, W) > 0$, which implies that at least one eigenvalue of $W_H$ is non-zero. □

Having established the convergence of the spectrum of $\mathcal{B}_H(G_n)$, it remains to derive the asymptotic distribution of $\Delta_2(H, G_n)$, and hence $\Gamma(H, G_n)$. This follows by 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 [6, Lemma 7.3].

**Lemma 4.5.** [6, Lemma 7.3] Let $Q_n = ((Q_n(i,j)))_{i,j \in |V(G_n)|}$ be a sequence of symmetric $|V(G_n)| \times |V(G_n)|$ matrices with zeros on the diagonal. If there exists constants for $\lambda_1 \geq \lambda_2 \geq \cdots$ such that $\lim_{n \to \infty} \text{tr}(Q_n^s) = \sum_{r=1}^{\infty} \lambda^s_r < \infty$, for every $s \geq 2$, then

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

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

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 [6, Theorem 1.4].

**Example 5.** (Monochromatic Edges) Let $G_n \Rightarrow W$ and $H = K_2$. Then $|\text{Aut}(H)| = 2$ 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 $\lambda_1(W) \geq \lambda_2(W) \geq \cdots$ are the eigenvalues of the operator $T_W : L_2[0,1] \to L_2[0,1]$, defined as $(T_W f)(x) = \int_0^1 W(x,y)f(y)dy$. Then Theorem 1.3 shows

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

where $\{\eta_r\}_{r \in \mathbb{N}}$ are independent $\chi^2_{c-1} - (c-1)$ random variables, as in [6, Theorem 1.4].

As before, Theorem 1.3 applies to convergent sequence of dense random graphs, when the limit in (1.5) hold 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 $G_n \Rightarrow W(p) = p$, the constant function $p$, and,
from (1.12), $W_H^{(p)}(x, y) = \frac{\Gamma(H)}{|Aut(H)|} p^{|E(H)|}$. It is easy to see that $W_H^{(p)}$ has only 1 non-zero eigenvalue

$$\lambda_1(W_H^{(p)}) = \frac{\Gamma(H)}{|Aut(H)|} p^{|E(H)|}.$$ 

Therefore, by Theorem 1.3,

$$\Gamma(H, K_n) \xrightarrow{D} \frac{\sigma_{H,p}}{c^{|V(H)|-1}} \cdot \left( \chi^2_{(c-1)} - (c-1) \right),$$

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

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 $G_n = K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$ and $H = K_{1,2}$. Then $|Aut(H)| = 2$, and $G_n \Rightarrow W = 1\{(x - \frac{1}{2}) (y - \frac{1}{2}) \leq 0\}$. This implies $d_W(x) = \frac{1}{2}$ for all $x \in [0, 1]$ and

$$W_{K_{1,2}}(x, y) = \frac{W(x, y) d_W(x) + W(x, y) d_W(y) + \int_{[0,1]} W(x, z_1) W(y, z_1) dz_1}{2}.$$

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

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

where $\eta_1$ and $\eta_2$ are independent $\chi^2_{(c-1)} - (c-1)$ 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 $G_n = K_{\lfloor \frac{n}{3} \rfloor, \lfloor \frac{n}{3} \rfloor, \lceil \frac{n}{3} \rceil}$ and $H = K_3$. Then $|Aut(H)| = 3$, and $G_n \Rightarrow W = 1\{(x, y) \in \mathcal{S}^c\}$, where $\mathcal{S} := [0, \frac{1}{3})^2 \cup \left[\frac{1}{3}, \frac{2}{3}\right]^2 \cup \left[\frac{2}{3}, 1\right]^2$. A direct computation gives that for all $(u, v) \in V(K_3)$, with $u \neq v$,

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

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

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

where $\eta_1, \eta_2, \eta_3$ are independent $\chi^2_{(c-1)} - (c-1)$ 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 Bin$(n, 1/c)$, and consequently

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

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 $D_n$ 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 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. In fact, in this case, there is no centering and scaling for which $T(H, G_n)$ has a non-degenerate limiting distribution).

References

[1] N. I. Akhiezer, The Classical Problem of Moments and Some Related Questions of Analysis, Oliver & Boyd, Edinburgh, 1965.

[2] A. D. Barbour, L. Holst, and S. Janson, Poisson Approximations, Oxford University Press, Oxford, 1992.

[3] A. Basak and S. Mukherjee, Universality of the mean-field for the Potts model, Probability Theory and Related Fields, to appear, 2017.

[4] B. Bhattacharya, Collision times in multicolor urn models and sequential graph coloring with applications to discrete lograrithms, Annals of Applied Probability, Vol. 26 (6), 3286–3318, 2016.

[5] B. Bhattacharya and S. Mukherjee, Inference in Ising models, Bernoulli, to appear, 2017.

[6] B. Bhattacharya, P. Diaconis, and S. Mukherjee, Universal Poisson and Normal limit theorems in graph coloring problems with connections to extremal combinatorics, Annals of Applied Probability, Vol. 27 (1), 337–394, 2017.

[7] 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, Vol. 219, 1801–1851, 2009.

[8] 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, Vol. 176, 151–219, 2012.

[9] A. Cerquetti and S. Fortini, A Poisson approximation for colored graphs under exchangeability, Sankhya: The Indian Journal of Statistics Vol. 68(2), 183–197, 2006.

[10] S. Chatterjee, Stein’s method and applications, UC Berkeley STAT C206A Lecture Notes, 2007. (https://statweb.stanford.edu/~souravc/Lecture6.pdf)

[11] S. Chatterjee, P. Diaconis, and E. Meckes, Exchangeable pairs and Poisson approximation, Electron. Encyclopedia Probab., 2004.

[12] A. DasGupta, The matching, birthday and the strong birthday problem: a contemporary review, J. Statist. Plann. Inferrence, Vol. 130, 377–389, 2005.

[13] P. Diaconis and F. Mosteller, Methods for studying coincidences, Journal of the American Statistical Association, Vol. 84(408), 853–861, 1989.

[14] F. M. Dong, K. M. Koh, and K. L. Teo, Chromatic polynomials and chromaticity of graphs, World Scientific Publishing Company, 2005.

[15] X. Fang, A universal error bound in the CLT for counting monochromatic edges in uniformly colored graphs, Electronic Communications in Probability, Vol. 20, Article 21, 1–6, 2015.

[16] J. H. Friedman and L. C. Rafsky, Multivariate generalizations of the Wolfowitz and Smirnov two-sample tests, Annals of Statistics, Vol. 7, 697–717, 1979.

[17] S. D. Galbraith, M. Holmes, A non-uniform birthday problem with applications to discrete logarithms, Discrete Applied Mathematics, Vol. 160 (10-11), 1547–1560, 2012.

[18] T. R. Jensen and B. Toft, Graph Coloring Problems, Wiley-Interscience Series in Discrete Mathematics and Optimization, 1995.

[19] T. R. Jensen and B. Toft, Unsolved Graph Coloring Problems, In L. Beineke & R. Wilson (Eds.), Topics in Chromatic Graph Theory, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 327–357, 2015.

[20] L. Lovász, Large Networks and Graph Limits, Colloquium Publications, Vol. 60, 2012.

[21] M. Nandi and D. R. Stinson, Multicollision attacks on some generalized sequential hash functions, IEEE Transactions on Information Theory, Vol. 53 (2), 759–767, 2007.

Department of Statistics, University of Pennsylvania, Philadelphia, USA, bhaswar@wharton.upenn.edu

Department of Statistics, Columbia University, New York, USA, sm3949@columbia.edu