RESEARCH ARTICLE

Limit theorems for monochromatic stars

Bhaswar B. Bhattacharya[1] | **Sumit Mukherjee[2]**

[1]Department of Statistics, University of Pennsylvania, Philadelphia, Pennsylvania,
[2]Department of Statistics, Columbia University, New York, New York

Correspondence
Bhaswar B. Bhattacharya, Department of Statistics, University of Pennsylvania, Philadelphia, PA.
Email: bhaswar@wharton.upenn.edu

Funding information
This research was supported by the National Science Foundation, NSF grant; DMS-1712037

Abstract

Let $T(K_{1,r}, G_n)$ be the number of monochromatic copies of the r-star $K_{1,r}$ in a uniformly random coloring of the vertices of the graph G_n. In this paper we provide a complete characterization of the limiting distribution of $T(K_{1,r}, G_n)$, in the regime where $\mathbb{E}(T(K_{1,r}, G_n))$ is bounded, for any growing sequence of graphs G_n. The asymptotic distribution is a sum of mutually independent components, each term of which is a polynomial of a single Poisson random variable of degree at most r. Conversely, any limiting distribution of $T(K_{1,r}, G_n)$ has a representation of this form. Examples and connections to the birthday problem are discussed.

KEYWORDS

combinatorial probability, graph coloring, limit theorems, subgraph counts

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:

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

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

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

the number of monochromatic edges in G_n.

The statistic (1.2) arises in several contexts, for example, as the Hamiltonian of the Ising/Potts models on G_n [2], in nonparametric two-sample tests [15], and the discrete logarithm problem [16]. Moreover, the asymptotics of $T(K_2, G_n)$ is often useful in the study of coincidences [12] as a generalization of the birthday paradox [1, 10–12]: If G_n is a friendship-network graph colored uniformly with $c_n = 365$ colors (corresponding to birthdays), then two friends will have the same birthday whenever the corresponding edge in the graph G_n is monochromatic.[1] Therefore, $\mathbb{P}(T(K_2, G_n) > 0)$ is the probability that there are two friends with the same birthday. 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 [13, 17, 20].

It is well known that the limiting distribution of $T(K_2, G_n)$ exhibits a universality, that is, $T(K_2, G_n) \xrightarrow{D} \text{Pois}(\lambda)$, whenever $\mathbb{E}(T(K_2, G_n)) = \frac{|E(G_n)|}{c_n} \to \lambda$, *for any graph sequence* G_n, as soon as $c_n \to \infty$. This was shown by Barbour et al. [1, Theorem 5.G], using the Stein's method for Poisson approximation, for any sequence of deterministic graphs. Recently, Bhattacharya et al. [4, Theorem 1.1] gave a new proof of this result based on the method of moments, which illustrates interesting connections to extremal combinatorics.

For a general graph H, define $T(H, G_n)$ to be the number of monochromatic copies of H in G_n, where the vertices of G_n are colored uniformly at random with c_n colors as in (1.1). Conditions under which $T(H, G_n)$ is asymptotically Poisson are easy to derive using Stein's method based on dependency graphs [7, 9]. However, the class of possible limiting distributions of $T(H, G_n)$, for a general graph H in the regime where $\mathbb{E}(T(H, G_n)) = O(1)$, can be extremely diverse (including mixture and polynomials in Poissons [4]), and there is no natural universality, as in the case of edges. Recently, Bhattacharya et al. [5] proved the following second-moment phenomenon for the asymptotic Poisson distribution of $T(H, G_n)$, for any connected graph H: $T(H, G_n)$ converges to $\text{Pois}(\lambda)$ whenever $\mathbb{E}T(H, G_n) \to \lambda$ and $\text{Var } T(H, G_n) \to \lambda$. Moreover, for any graph H, $T(H, G_n)$ converges to linear combination of independent Poisson variables, when G_n is a converging sequence of dense graphs [6].

However, there is no description of the set of possible limits of $T(H, G_n)$, other than the case of monochromatic edges ($H = K_2$) or dense graphs G_n (where the limits are Poisson or a linear combination of independent Poissons, respectively). In this paper, we consider the case of the r-star ($H = K_{1,r}$). This arises as a generalization of the birthday problem, for example, with $r = 2$ and a friendship network G_n, $T(K_{1,2}, G_n)$ counts the number of triples with the same birthday where someone is friends with the other two. This is especially relevant when G_n has a few influential nodes which have many friends ("superstar" vertices [3]), and we wish to count the number of triple birthday matches with a superstar. Related statistics also appear as the Hamiltonian in exponential random graph models (ERGMs) [8, 18, 19], which capture the more realistic scenario of dependent edges in social networks, as opposed to the Erdős-Rényi random graph model where the edges are mutually independent.

In this paper we identify the set of all possible limiting distributions of $T(K_{1,r}, G_n)$, for *any graph sequence* G_n. We show that the asymptotic distribution of $T(K_{1,r}, G_n)$ is a sum of mutually independent components, each term of which is a polynomial of a single Poisson random variable of degree at most r, and, conversely, any limiting distribution of $T(K_{1,r}, G_n)$ has this form.

1.1 | Limiting distribution for monochromatic r-stars

Let G_n be a simple graph with vertex set $V(G_n)$ and edge set $E(G_n)$. For a fixed graph H, denote by $N(H, G_n)$ the number of isomorphic copies of H in G_n. Note that $N(K_{1,r}, G_n) = \sum_{v \in V(G_n)} \binom{d_v}{r}$, where d_v is the degree of the vertex $v \in V(G_n)$.

---

[1]When the underlying graph $G_n = K_n$ is the complete graph K_n on n vertices, this reduces to the classical birthday problem.

Now, suppose G_n is colored with c_n colors as in (1.1). If X_v denotes the color of vertex $v \in V(G_n)$, then the number of monochromatic copies of $K_{1,r}$ in G_n is

$$T(K_{1,r}, G_n) := \sum_{v=1}^{|V(G_n)|} \sum_{\boldsymbol{u} \in \binom{V(G_n)}{r}} a_v(\boldsymbol{u}, G_n)\mathbf{1}\{X_v = X_{\boldsymbol{u}}\}, \tag{1.3}$$

where

- $\binom{V(G_n)}{r}$ is the collection of r-element subsets of G_n;
- $a_v(\boldsymbol{u}, G_n) = \prod_{s=1}^{r} a_{vu_s}(G_n)$, for $v \in V(G_n)$ and $\boldsymbol{u} = \{u_1, u_2, \ldots, u_r\} \in \binom{V(G_n)}{r}$;
- $\mathbf{1}\{X_v = X_{\boldsymbol{u}}\} := \mathbf{1}\{X_v = X_{u_1} = \cdots = X_{u_r}\}$, for $v \in V(G_n)$ and $\boldsymbol{u} \in \binom{V(G_n)}{r}$, as above.

Note that

$$\mathbb{E}(T(K_{1,r}, G_n)) = \frac{1}{c_n^r} \sum_{v=1}^{|V(G_n)|} \sum_{\boldsymbol{u} \in \binom{V(G_n)}{r}} a_v(\boldsymbol{u}, G_n) = \frac{1}{c_n^r} N(K_{1,r}, G_n).$$

It is known that the limiting behavior of $T(K_{1,r}, G_n)$ is governed by its expectation:

Proposition 1.1 [5, Lemma 3.1] *Let $\{G_n\}_{n\geq 1}$ be a sequence of deterministic graphs colored uniformly with c_n colors as in (1.1). Then*

$$T(K_{1,r}, G_n) \xrightarrow{P} \begin{cases} 0 & \text{if } \lim_{n\to\infty} \mathbb{E}(T(K_{1,r}, G_n)) = 0, \\ \infty & \text{if } \lim_{n\to\infty} \mathbb{E}(T(K_{1,r}, G_n)) = \infty. \end{cases}$$

Therefore, the most interesting regime is where $\mathbb{E}(T(K_{1,r}, G_n)) = \Theta(1)$,[2] that is, $c_n \to \infty$ such that

$$\mathbb{E}(T(K_{1,r}, G_n)) = \frac{N(K_{1,r}, G_n)}{c_n^r} = \frac{1}{c_n^r} \sum_{v \in V(G_n)} \binom{d_v}{r} = \Theta(1). \tag{1.4}$$

Theorem 1.2 *Let $\{G_n\}_{n\geq 1}$ be a sequence of graphs colored uniformly with c_n colors with $c_n \to \infty$, as in (1.1).*

(a) Assume that the following hold:

(1) For every $k \in [1, r+1]$, there exists $\lambda_k \geq 0$ such that

$$\lim_{n\to\infty} \frac{\sum_{F \in \mathscr{C}_{r,k}} N_{\text{ind}}(F, G_n)}{c_n^r} = \lambda_k, \tag{1.5}$$

where $N_{\text{ind}}(F, G_n)$ is the number of induced copies of F in G_n and $\mathscr{C}_{r,k} := \{F \supseteq K_{1,r} : |V(F)| = r+1 \text{ and } N(K_{1,r}, F) = k\}$.

(2) Let $d_{(1)} \geq d_{(2)} \geq \cdots \geq d_{(|V(G_n)|)}$ be the degrees of the vertices in G_n arranged in nonincreasing order, such that

$$\lim_{n\to\infty} \frac{d_{(v)}}{c_n} = \theta_v, \tag{1.6}$$

for each $v \in V(G_n)$ fixed.

---

[2]For two nonnegative sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$, $a_n = \Theta(b_n)$ means that there exist positive constants C_1, C_2, such that $C_1 b_n \leq a_n \leq C_2 b_n$, for all n large enough.

Then

$$T(K_{1,r}, G_n) \to \sum_{v=1}^{\infty} \binom{T_v}{r} + \sum_{k=1}^{r+1} k Z_k, \tag{1.7}$$

where the convergence is in distribution and in all moments, and

- $T_1, T_2, \ldots$, *are independent* $\text{Pois}(\theta_1), \text{Pois}(\theta_2), \ldots$, *respectively;*
- $\lambda_1 - \frac{1}{r!} \sum_{u=1}^{\infty} \theta_u^r \geq 0$, *and* $Z_1, Z_2, \ldots, Z_{r+1}$ *are independent* $\text{Pois}(\lambda_1 - \frac{1}{r!} \sum_{u=1}^{\infty} \theta_u^r)$, $\text{Pois}(\lambda_2), \ldots \text{Pois}(\lambda_{r+1})$, *respectively;*
- *the collections* $\{T_k, k \geq 1\}$ *and* $\{Z_k, 1 \leq k \leq r+1\}$ *are independent.*

(b) *Conversely, if* $T(K_{1,r}, G_n)$ *converges in distribution, then the limit is necessarily of the form as in the RHS of (1.7), for some nonnegative constants* $\theta_1 \geq \theta_2 \geq \cdots$, *and* $\{\lambda_k, 1 \leq k \leq r+1\}$.

This result gives a complete characterization of the limiting distribution of $T(K_{1,r}, G_n)$, in the regime where $\mathbb{E}(T(K_{1,r}, G_n)) = \Theta(1)$ (in fact, under the assumptions of the theorem $\mathbb{E}(T(K_{1,r}, G_n)) \to \sum_{k=1}^{r+1} k\lambda_k$). Note that the limit in (1.7) has two components:

- a *nonlinear part* $\sum_{v=1}^{\infty} \binom{T_v}{r}$ which corresponds to the number of monochromatic $K_{1,r}$ in G_n with central vertex of "high" degree, that is, the vertices of degree $\Theta(c_n)$; and
- a *linear part* $\sum_{k=1}^{r+1} k Z_k$ which is the number of monochromatic $K_{1,r}$ from the "low" degree vertices, that is, degree $o(c_n)$;

and, perhaps interestingly, the linear and the nonlinear parts are asymptotically independent. The proof is given in Section 2. It involves decomposing the graph based on the degree of the vertices, and then using moment comparisons, to establish independence and compute the limiting distribution.

Remark 1.1. An easy sufficient condition for (1.5) is the convergence of $\frac{1}{c_n^{|V(F)|-1}} N_{\text{ind}}(K_{1,r}, G_n)$ for *every* super-graph F of $K_{1,r}$ with $|V(F)| = r+1$. However, condition (1.5) does not require the convergence for every such graph, and is applicable to more general examples, as described below: Define a sequence of graphs G_n as follows:

$$G_n = \begin{cases} \text{disjoint union of } n \text{ isomorphic copies of the 3-star } K_{1,3} & \text{if } n \text{ is odd} \\ \text{disjoint union of } n \text{ isomorphic copies of the } (3, 1)\text{-tadpole } \Delta_+ & \text{if } n \text{ is even,} \end{cases}$$

where the $(3, 1)$-tadpole is the graph obtained by joining a triangle and a single vertex with a bridge. Now, choosing $c_n = \lfloor n^{1/3} \rfloor$, gives $\mathbb{E}(T(K_{1,3}, G_n)) \to 1$. In this case,

$$\frac{\sum_{F \in \mathscr{C}_{H,1}} N_{\text{ind}}(F, G_n)}{c_n^3} = \frac{N_{\text{ind}}(K_{1,3}, G_n) + N_{\text{ind}}(\Delta_+, G_n)}{c_n^3} \to 1,$$

and $\frac{1}{c_n^3} \sum_{F \in \mathscr{C}_{H,4}} N_{\text{ind}}(F, G_n) = \frac{1}{c_n^3} \sum_{F \in \mathscr{C}_{H,4}} N_{\text{ind}}(F, G_n) = \frac{1}{c_n^3} \sum_{F \in \mathscr{C}_{H,2}} N_{\text{ind}}(F, G_n) = 0$. Therefore, Theorem 1.2 implies that $T(K_{1,3}, G_n) \xrightarrow{D} \text{Pois}(1)$ (which can also be directly verified, because, in this case, $T(K_{1,3}, G_n)$ is a sum of independent $\text{Ber}(\frac{1}{c_n^3})$ variables). However, it is easy to see that individually both $\frac{1}{c_n^3} N_{\text{ind}}(K_{1,3}, G_n)$ and $\frac{1}{c_n^3} N_{\text{ind}}(\Delta_+, G_n)$ are nonconvergent.

The limit in (1.7) simplifies when the graph G_n has no vertices of high degree. The following corollary is a consequence of Theorem 1.2.

Corollary 1.3 *Let $\{G_n\}_{n \geq 1}$ be a sequence of deterministic graphs. Then the following are equivalent.*

(a) Condition (1.5) and $\lim_{n \to \infty} \frac{\Delta(G_n)}{c_n} = 0$, where $\Delta(G_n) := \max_{v \in V(G_n)} d_v$.

(b) $T(K_{1,r}, G_n) \overset{D}{\to} \sum_{k=1}^{r+1} k Z_k$, where $Z_1, \dots, Z_{r+1}$ are independent $\mathrm{Pois}(\lambda_1), \dots \mathrm{Pois}(\lambda_{r+1})$, respectively.

The proof of the corollary is given in Section 2.6. Applications of this corollary and Theorem 1.2 are discussed in Section 3. In Section 4 we discuss open problems and directions for future research.

2 | PROOFS OF THEOREM 1.2 AND COROLLARY 1.3

The proof of Theorem 1.2 has four main steps:

(1) Decomposing G_n into the "high"-degree and "low"-degree vertices, and showing that the resulting error term vanishes (Section 2.1).
(2) Showing that the contributions from the "high"-degree and "low"-degree vertices are asymptotically independent in moments (Section 2.2).
(3) Computing the limiting distribution of the number of monochromatic r-stars with central vertex at one of the "high"-degree vertices, which gives the nonlinear term in (1.7) (Section 2.3).
(4) Computing the limiting distribution of the number of monochromatic r-stars from the "low"-degree vertices, which gives the linear combination of independent Poisson variables in (1.7) (Section 2.4).

The proof of Theorem 1.2 can be easily completed by combining the above steps (Section 2.5). The proof of Corollary 1.3 is given in Section 2.6.

2.1 | Decomposing G_n

Before proceeding, we recall some standard asymptotic notation. For two nonnegative sequences $(a_n)_{n \geq 1}$ and $(b_n)_{n \geq 1}$, $a_n \lesssim b_n$ means $a_n = O(b_n)$, and $a_n \sim b_n$ means $a_n = (1 + o(1))b_n$. We will use subscripts in the above notation, for example, $O_\square(\cdot)$, $\lesssim_\square$ to denote that the hidden constants may depend on the subscripted parameters.

To begin with, note that the number of r-stars in G_n remains unchanged if all edges (u, v) in G_n such that $\max\{d_u, d_v\} \leq r - 1$ are dropped. Hence, without loss of generality, assume that $\max\{d_u, d_v\} \geq r$, for all edges $(u, v) \in G_n$. This ensures that $N(K_{1,r}, G_n) = \sum_{v \in V(G_n)} \binom{d_v}{r}$ has the same order as $\sum_{v \in V(G_n)} d_v^r$ as shown below:

Observation 2.1 *If $\max\{d_u, d_v\} \geq r$, for all edges $(u, v) \in G_n$, then assumption (1.4) implies*

$$\sum_{v \in V(G_n)} d_v^r = \Theta(c_n^r). \tag{2.1}$$

Proof In this case, the following inequality holds

$$\frac{1}{2} \sum_{v \in V(G_n)} d_v \leq \sum_{v \in V(G_n)} d_v \mathbf{1}\{d_v \geq r\}. \tag{2.2}$$

To see this note that if an edge $(u, v) \in E(G_n)$ has $\min\{d_u, d_v\} \geq r$, then that edge is counted two times in the RHS above, and an edge $(u, v) \in E(G_n)$ which has $\min\{d_u, d_v\} \leq r - 1$ (but

$\max\{d_u, d_v\} \geq r$) is counted once in the RHS, whereas every edge of $E(G_n)$ is counted twice in the LHS.

Then

$$
\begin{aligned}
\sum_{v \in V(G_n)} d_v^r &= \sum_{v \in V(G_n)} d_v^r \mathbf{1}\{d_v < r\} + \sum_{v \in V(G_n)} d_v^r \{d_v \geq r\} \\
&\leq (r-1)^{r-1} \sum_{v \in V(G_n)} d_v + r^r \sum_{v \in V(G_n)} \binom{d_v}{r} \\
&\leq 2r^{r-1} \sum_{v \in V(G_n)} d_v \mathbf{1}\{d_v \geq r\} + r^r \sum_{v \in V(G_n)} \binom{d_v}{r} \qquad \text{(using (2.2))} \\
&\leq 2r^r \sum_{v \in V(G_n)} \binom{d_v}{r} + r^r \sum_{v \in V(G_n)} \binom{d_v}{r} = 3r^r \sum_{v \in V(G_n)} \binom{d_v}{r},
\end{aligned}
$$

from which the desired conclusion follows on using (1.4). ∎

Throughout the rest of this section, we will thus assume, that $\max\{d_u, d_v\} \geq r$, for all edges $(u, v) \in G_n$ and, hence, (1.4) implies (2.1). Note that (2.1) implies

$$
\Delta(G_n) := \max_{v \in V(G_n)} d_v = O(c_n).
$$

In fact, using (2.1) it can be shown that there are not too many vertices $v \in V(G_n)$ with $d_v = \Theta(c_n)$. To this end, for $\theta_1 \geq \theta_2 \geq \cdots$ as in (1.6), let $\|\theta\|_\infty := \max_{u \in \mathbb{N}} \theta_u$. Now, we have the following definition:

Definition 2.1. If $\|\theta\|_\infty > 0$, fix $\varepsilon \in (0, \|\theta\|_\infty)$ such that $\varepsilon \neq \theta_u$ for any $u \in \mathbb{N}$. (This can be done, as the set $\{\theta_u, u \in \mathbb{N}\}$ is countable.) A vertex $v \in V(G_n)$ is said to be ε-big if $d_v \geq \varepsilon c_n$. Denote the subset of ε-big vertices by $V_\varepsilon(G_n)$. Let $\eta = \eta(\varepsilon) \in \mathbb{N}$ be such that $\theta_\eta > \varepsilon > \theta_{\eta+1}$ (which exists, as $\theta_u \downarrow 0$ when $u \to \infty$).

The following lemma is an easy consequence of (2.1) and the above definition.

Lemma 2.1 *Assume (2.1) holds and $\theta_1 \geq \theta_2 \geq \cdots$ as in (1.6). Then the following hold for all large n:*

(a) *If $\|\theta\|_\infty = 0$, then the set ε-big vertices $|V_\varepsilon(G_n)|$ is empty.*

(b) *If $\|\theta\|_\infty > 0$, then for all $\varepsilon \in (0, \|\theta\|_\infty)$ such that $\varepsilon \neq \theta_u$ the number of ε-big vertices $|V_\varepsilon(G_n)|$ equals η (as defined in Definition 2.1), which does not depend on n.*

Proof If $\|\theta\|_\infty = 0$ (which is equivalent to $\max_{v \in G_n} d_v = o(c_n)$), given $\varepsilon > 0$ we have $\max_{v \in G_n} d_v < \varepsilon c_n$ for all n large. This means $V_\varepsilon(G_n)$ is empty, which proves (a).

For (b), for all n large enough we have $d_{(\eta+1)} < \varepsilon c_n < d_{(\eta)}$, hence $|V_\varepsilon(G_n)| = \eta$, as desired. ∎

Hereafter, we proceed with the case $\|\theta\|_\infty > 0$, which implies by Lemma 2.1 that $V_\varepsilon(G_n)$ is not empty, and has cardinality $\eta(\varepsilon)$ free of n. We will return to the case $\|\theta\|_\infty = 0$ at the end.

Definition 2.2. Denote by $G_{n,\varepsilon}$ the subgraph of G_n obtained by removing the edges between the ε-big vertices, and let $T(K_{1,r}, G_{n,\varepsilon})$ be the number of monochromatic r-stars in $G_{n,\varepsilon}$.

The following lemma shows that removing the edges between the ε-big vertices of G_n does not change the number of monochromatic r-stars in G_n, in the limit.

Lemma 2.2 *Assume (1.4) holds. Then for every fixed $\varepsilon > 0$, as $n \to \infty$,*

$$N(F, G_n) - N(F, G_{n,\varepsilon}) = o(c_n^r) \quad and \quad N_{\text{ind}}(F, G_n) - N_{\text{ind}}(F, G_{n,\varepsilon}) = o(c_n^r),$$

for all $F \supseteq K_{1,r}$ with $|V(F)| = r + 1$. Consequently,

$$\lim_{n \to \infty} \mathbb{E}|T(K_{1,r}, G_n) - T(K_{1,r}, G_{n,\varepsilon})| \to 0.$$

Proof If a graph $F \supseteq K_{1,r}$ with $|V(F)| = r + 1$ is a subgraph of G_n, but not a subgraph of $G_{n,\varepsilon}$, then it must have at least one edge with both end-points in $V_\varepsilon(G_n)$. Choosing this edge in $|V_\varepsilon(G_n)|^2$ ways and the remaining $r-1$ vertices in $O(c_n^{r-1})$ ways (since the maximum degree $\Delta(G_n) = O(c_n)$), it follows that

$$N(F, G_n) - N(F, G_{n,\varepsilon}) = O(c_n^{r-1}|V_\varepsilon(G_n)|^2) = o(c_n^r),$$

as $n \to \infty$, since by Lemma 2.1 $|V_\varepsilon(G_n)| = O_\varepsilon(1)$. As the number of induced copies of F in G_n which are not in $G_{n,\varepsilon}$, is bounded by the total number of copies of F in $G_{n,\varepsilon}$ which are not in G_n, the result on induced copies follows.

In particular,

$$\mathbb{E}|T(K_{1,r}, G_n) - T(K_{1,r}, G_{n,\varepsilon})| \lesssim \frac{1}{c_n^r}(c_n^{r-1}|V_\varepsilon(G_n)|^2) = \frac{1}{c_n}|V_\varepsilon(G_n)|^2 \to 0,$$

as $c_n \to \infty$. ∎

We now decompose the graph $G_{n,\varepsilon}$ based on the degree of the vertices as follows:

- Let $G_{n,\varepsilon}^+$ be the subgraph of $G_{n,\varepsilon}$ formed by the ε-big vertices and the edges incident on them. More formally, it has vertex set $V_\varepsilon(G_n) \bigcup N_{G_{n,\varepsilon}}(V_\varepsilon(G_n))$, where $N_{G_{n,\varepsilon}}(V_\varepsilon(G_n))$ is neighborhood of $V_\varepsilon(G_n)$ in $G_{n,\varepsilon}$,[3] and edge set $\{(u, v) \in E(G_{n,\varepsilon}) : v \in V_\varepsilon(G_n)\}$. Note that by construction $G_{n,\varepsilon}^+$ is a bipartite graph.
- Let $G_{n,\varepsilon}^-$ denote the induced subgraph of $G_{n,\varepsilon}$ with vertex set $V(G_n)\backslash V_\varepsilon(G_n)$.

The decomposition of the graph $G_{n,\varepsilon}$ is illustrated in Figure 1. Note that $G_{n,\varepsilon}^+$ and $G_{n,\varepsilon}^-$ have common vertices (the black vertices in Figure 1), but no common edges, and consequently no common r-stars. This implies

$$T(K_{1,r}, G_{n,\varepsilon}) = T_+(K_{1,r}, G_{n,\varepsilon}^+) + T(K_{1,r}, G_{n,\varepsilon}^-) + R(K_{1,r}, G_{n,\varepsilon}),$$

where $T(K_{1,r}, G_{n,\varepsilon}^-)$ is the number of monochromatic r-stars in $G_{n,\varepsilon}^-$; and (recalling the definition of $a_v(\boldsymbol{u}, G_n)$ from (1.3))

$$T_+(K_{1,r}, G_{n,\varepsilon}^+) := \sum_{v=1}^{|V_\varepsilon(G_n)|} \sum_{\boldsymbol{u} \in \binom{V(G_n)}{r}} a_v(\boldsymbol{u}, G_n)\mathbf{1}\{X_v = X_{\boldsymbol{u}}\}, \tag{2.3}$$

counts the number of monochromatic r-stars in $G_{n,\varepsilon}^+$ with central vertex in $V_\varepsilon(G_n)$;[4] and the *remainder term*

$$R(K_{1,r}, G_{n,\varepsilon}) := \sum_{v \notin V_\varepsilon(G_n)} \sum_{u_1 \in V_\varepsilon(G_n)} \sum_{\boldsymbol{u} \in \binom{V(G_n)}{r-1}} a_{vu_1}(G_n)a_v(\boldsymbol{u}, G_n)\mathbf{1}\{X_v = X_{u_1} = X_{\boldsymbol{u}}\}. \tag{2.4}$$

---

[3]For a graph $H = (V(H), E(H))$ and $S \subseteq V(H)$, the *neighborhood* of S in H is $N_H(S) = \{v \in V(H) : \exists\, u \in S \text{ such that } (u, v) \in E(H)\}$.

[4]Note that $T_+(K_{1,r}, G_{n,\varepsilon}^+)$ is not the number of r-stars in $G_{n,\varepsilon}^+$: It does not include the r-stars in $G_{n,\varepsilon}^+$ with central vertex in $N_{G_{n,\varepsilon}}(V_\varepsilon(G_n))$ (the black vertices in Figure 1). Instead, these r-stars are included in the remainder term $R(K_{1,r}, G_{n,\varepsilon})$.
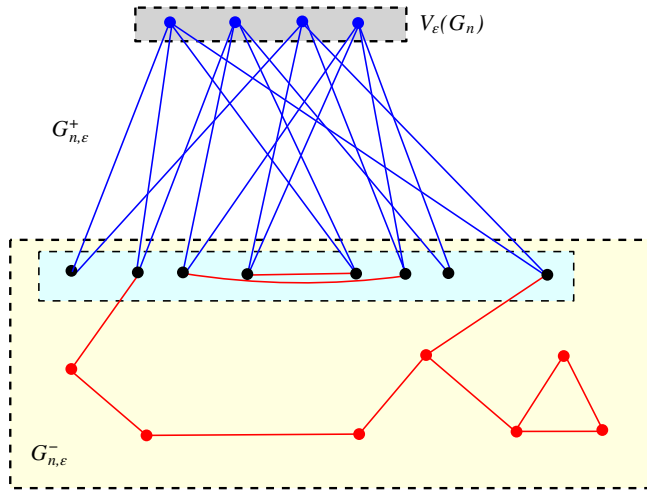
FIGURE 1 The decomposition of $G_{n,\varepsilon}$: the graph formed by the blue edges is $G_{n,\varepsilon}^+$ and the graph formed by the red edges is $G_{n,\varepsilon}^-$. Note that the black vertices belong to both $G_{n,\varepsilon}^+$ and $G_{n,\varepsilon}^-$ [Color figure can be viewed at wileyonlinelibrary.com]

The following lemma shows that the remainder term goes to zero in expectation, and therefore, in probability.

Lemma 2.3   *Let $R(K_{1,r}, G_{n,\varepsilon})$ be as defined above in (2.4). Under the assumptions of Theorem 1.2,*

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \mathbb{E}(R(K_{1,r}, G_{n,\varepsilon})) = 0.$$

Proof   Note that for $v \notin V_\varepsilon(G_n)$

$$\sum_{\mathbf{u} \in \binom{V(G_n)}{r-1}} a_v(\mathbf{u}, G_n) \le d_v^{r-1} \le (\varepsilon c_n)^{r-1}. \tag{2.5}$$

Moreover $\sum_{v \notin V_\varepsilon(G_n)} a_{vu_1}(G_{n,\varepsilon}) \le d_{u_1}$. Then, using (2.4), for any $M > 0$,

$$\mathbb{E}(R(K_{1,r}, G_{n,\varepsilon})) \le \frac{(\varepsilon c_n)^{r-1}}{c_n^r} \sum_{u_1 \in V_\varepsilon(G_n)} d_{u_1}$$

$$= \frac{\varepsilon^{r-1}}{c_n} \left(\sum_{u : \varepsilon c_n \le d_u < M\varepsilon c_n} d_u + \sum_{u : d_u \ge M\varepsilon c_n} d_u \right)$$

$$\le \frac{1}{c_n^r} \sum_{u : \varepsilon c_n \le d_u < M\varepsilon c_n} d_u^r + \frac{1}{M^{r-1} c_n^r} \sum_{u \in V(G_n)} d_u^r. \tag{2.6}$$

Since $\limsup_{n \to \infty} \frac{1}{c_n^r} \sum_{u \in V(G_n)} d_u^r < \infty$ (from Observation 2.1), the second term in the RHS of (2.6) converges to 0 on letting $n \to \infty$ followed by $M \to \infty$.

Next, recall from Lemma 2.1 that for all ε small enough there exists $\eta = \eta(\varepsilon)$ such that $\theta_{\eta+1} < \varepsilon < \theta_\eta$. Thus, for all n large enough, $d_{\eta+1} < \varepsilon c_n < d_\eta$, and as $n \to \infty$, the first term in the RHS

above becomes

$$\limsup_{n \to \infty} \sum_{u=1}^{\eta} \frac{d_u^r}{c_n^r} \mathbf{1}\{d_u < M\varepsilon c_n\} \leq \sum_{u=1}^{\eta} \theta_u^r \mathbf{1}\{\theta_u \leq M\varepsilon\} = \sum_{u=1}^{\infty} \theta_u^r \mathbf{1}\{\varepsilon \leq \theta_u \leq M\varepsilon\},$$

which converges to 0 on letting $\varepsilon \to 0$, by using DCT along with the fact that $\sum_{u=1}^{\infty} \theta_u^r \leq \limsup_{n \to \infty} \frac{1}{c_n^r} \sum_{u \in V(G_n)} d_u^r < \infty$ (by Fatou's lemma). ∎

Combining Lemmas 2.2 and 2.3 it follows that

$$
\begin{aligned}
T(K_{1,r}, G_n) &= T(K_{1,r}, G_{n,\varepsilon}) + o_P(1) \\
&= T_+(K_{1,r}, G_{n,\varepsilon}^+) + T(K_{1,r}, G_{n,\varepsilon}^-) + o_P(1).
\end{aligned}
\tag{2.7}
$$

Here the $o_P(1)$ term converges to 0 in probability as $n \to \infty$ followed by $\varepsilon \to 0$. Therefore, the limiting distribution of the $T(K_{1,r}, G_n)$ under the double limit is the same as that of $T_+(K_{1,r}, G_{n,\varepsilon}^+) + T(K_{1,r}, G_{n,\varepsilon}^-)$.

2.2 | Independence in moments of the contributions from $G_{n,\varepsilon}^+$ and $G_{n,\varepsilon}^-$

In this section we show that the number of monochromatic $K_{1,r}$ coming from $G_{n,\varepsilon}^+$ and $G_{n,\varepsilon}^-$ is asymptotically independent in moments. Without loss of generality, assume the vertices in $V(G_n)$ are labeled $1, 2, \ldots, |V(G_n)|$ such that $d_1 \geq d_2 \geq \cdots \geq d_{|V(G_n)|}$, and $\eta = \eta(\varepsilon)$ be as in Definition 2.1 such that $\theta_{\eta+1} < \varepsilon < \theta_\eta$. Then, by Definition (2.3),

$$T_+(K_{1,r}, G_{n,\varepsilon}^+) = \sum_{v=1}^{\eta} \binom{T_{G_{n,\varepsilon}^+}(v)}{r}, \quad \text{where} \quad T_{G_{n,\varepsilon}^+}(v) := \sum_{u \in V(G_{n,\varepsilon})} a_{uv}(G_n) \mathbf{1}\{X_u = X_v\}, \tag{2.8}$$

is the number of monochromatic r-stars in $G_{n,\varepsilon}$, with central vertex $v \in V_\varepsilon(G_n)$.

Now let $\eta_0 := \lim_{\varepsilon \to 0} \eta(\varepsilon)$, and fix a finite positive integer $K \leq \eta_0$. Then, for $\varepsilon > 0$ small enough, $\eta(\varepsilon) \geq K$, and so $\{T_{G_{n,\varepsilon}^+}(v) : 1 \leq v \leq K\}$ are well defined. The following lemma shows that this collection and $T(K_{1,r}, G_{n,\varepsilon}^-)$ are asymptotically independent in the moments.

Lemma 2.4 *Assume (1.4) holds. Then for every finite $K \leq \eta_0$ and nonnegative integers $s, t_1, \ldots, t_K$,*

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \left| \mathbb{E}\left(T(K_{1,r}, G_{n,\varepsilon}^-)^s \prod_{v=1}^{K} T_{G_{n,\varepsilon}^+}(v)^{t_v} \right) - \mathbb{E} T(K_{1,r}, G_{n,\varepsilon}^-)^s \left(\mathbb{E} \prod_{v=1}^{K} T_{G_{n,\varepsilon}^+}(v)^{t_v} \right) \right| = 0. \tag{2.9}$$

Proof of Lemma 2.4

For any labeled subgraph H of G_n, define

$$\beta(H) := \mathbb{E} \prod_{(u,v) \in E(H)} \mathbf{1}\{X_u = X_v\} = \left(\frac{1}{c_n} \right)^{|V(H)| - \nu(H)}, \tag{2.10}$$

where $\nu(H)$ is the number of connected components of H. Note that the definition of $\beta(\cdot)$ is invariant to the labeling of H, and so, it extends to unlabeled graphs as well. Thus, without loss of generality, we will define $\beta(H)$ as in (2.10), for an unlabeled graph H as well.

Let $H_1 = (V(H_1), E(H_1))$ and $H_2 = (V(H_2), E(H_2))$ be two (labeled) subgraphs of G_n, that is, $V(H_1)$ and $V(H_2)$ are subsets of $V(G_n)$, which inherits the labeling induced by $V(G_n)$, and $E(H_1)$ and $E(H_2)$ are subsets of $E(H)$. Let $H_1 \bigcup H_2 = (V(H_1) \bigcup V(H_2), E(H_1) \bigcup E(H_2))$.

Lemma 2.5 *For any two finite graphs H_1 and H_2, $\beta\left(H_1 \bigcup H_2\right) \geq \beta(H_1)\beta(H_2)$, where $\beta(\cdot)$ is defined above in (2.10).*

Proof Denote by $F = H_1 \bigcup H_2$, and let $F_1, F_2, \ldots, F_{v(F)}$ be the connected components of F. Define

$$I_1 = \{s \in [v(F)] : V(F_s) \bigcap V(H_1) \neq \varnothing \text{ and } V(F_s) \bigcap V(H_2) = \varnothing\},$$
$$I_2 = \{s \in [v(F)] : V(F_s) \bigcap V(H_1) = \varnothing \text{ and } V(F_s) \bigcap V(H_2) \neq \varnothing\},$$
$$I_{12} = \{s \in [v(F)] : V(F_s) \bigcap V(H_1) \neq \varnothing \text{ and } V(F_s) \bigcap V(H_2) \neq \varnothing\}. \quad (2.11)$$

Fix $s \in I_{12}$, that is, $V(F_s) \bigcap V(H_1) \neq \varnothing$ and $V(F_s) \bigcap V(H_2) \neq \varnothing$. Then $F_s = F'_s \bigcup F''_s$, where

$$F'_s = (V(F_s) \bigcap V(H_1), E(F_s) \bigcap E(H_1)), \quad \text{and} \quad F''_s = (V(F_s) \bigcap V(H_2), E(F_s) \bigcap E(H_2)).$$

Let $F'_{s1}, F'_{s2} \ldots F'_{sa}$ be the connected components of F'_s and similarly, $F''_{s1}, F''_{s2} \ldots F''_{sb}$ be the connected components of F''_s, where $a = v(F'_s)$ and $b = v(F''_s)$. Construct a bipartite graph $B_s = (B'_s \bigcup B''_s, E(B_s))$, where $B'_s = \{F'_{s1}, F'_{s2}, \ldots F'_{sa}\}$ and $B''_s = \{F''_{s1}, F''_{s2}, \ldots F''_{sb}\}$ and there is any edge between F'_{sx} and F''_{sy} if and only if $V(F'_{sx}) \bigcap V(F''_{sy}) \neq \varnothing$, for $x \in [a]$ and $y \in [b]$. Note that $|V(F'_s) \bigcap V(F''_s)| \geq |E(B_s)|$, and since the graph F_s is connected, the graph B_s is also connected. Therefore,

$$|V(F'_s) \bigcap V(F''_s)| \geq |E(B_s)| \geq |V(B_s)| - 1 = v(F'_s) + v(F''_s) - 1,$$

This implies,

$$|V(F_s)| = |V(F'_s)| + |V(F''_s)| - |V(F'_s) \bigcap V(F''_s)| \leq |V(F'_s)| - v(F'_s) + |V(F''_s)| - v(F''_s) + 1.$$

Then, recalling (2.11), it follows that

$$\beta(H) = \prod_{s \in I_1} \beta(F_s) \prod_{s \in I_2} \beta(F_s) \prod_{s \in I_{12}} \beta(F_s)$$
$$= \prod_{s \in I_{12}} \left(\frac{1}{c_n}\right)^{|V(F_s)|-1} \prod_{s \in I_1} \beta(F_s) \prod_{s \in I_2} \beta(F_s)$$
$$\geq \left(\prod_{s \in I_{12}} \left(\frac{1}{c_n}\right)^{|V(F'_s)|-v(F'_s)} \prod_{s \in I_1} \beta(F_s)\right) \left(\prod_{s \in I_{12}} \left(\frac{1}{c_n}\right)^{|V(F''_s)|-v(F''_s)} \prod_{s \in I_2} \beta(F_s)\right)$$
$$= \beta(H_1)\beta(H_2),$$

completing the proof of the lemma. ∎

Now, recall the definitions of the graph $G^-_{n,\varepsilon}$ from Section 2.1, and note that

$$T(K_{1,r}, G^-_{n,\varepsilon}) = \sum_{u \in \mathscr{S}_r(G^-_{n,\varepsilon})} \mathbf{1}\{X_{=u}\}, \quad (2.12)$$

where

- $\mathcal{S}_r(G_{n,\varepsilon}^-)$ is the collection of ordered $(r+1)$-tuples $\boldsymbol{u} = (u_0, u_1, \ldots, u_r)$, such that $u_0, u_1, \ldots, u_r \in V(G_{n,\varepsilon}^-)$ are distinct and $(u_0, u_i) \in E(G_{n,\varepsilon}^-)$, for $i \in [1, r]$; and
- $\mathbf{1}\{X_{=\boldsymbol{u}}\} = \mathbf{1}\{X_{u_0} = X_{u_1} = \cdots = X_{u_r}\}$.

For any $u \in V(G_n)$, let $N_{G_{n,\varepsilon}^+}(u)$ be the neighborhood of u in $G_{n,\varepsilon}^+$. Index the vertices in $N_{G_{n,\varepsilon}^+}(u)$ as $\{b_1(v), b_2(v), \ldots b_{d_v^+}(v)\}$, where d_v^+ is the degree of the vertex v in $G_{n,\varepsilon}^+$. Let

$$\Gamma = \prod_{v=1}^{K} N_{G_{n,\varepsilon}^+}(v)^{t_v} \times \mathcal{S}_r(G_{n,\varepsilon}^-)^s$$

denote the collection of vertices $\{b_j(v), 1 \le j \le t_v, 1 \le v \le K\}$ and s ordered $(r+1)$-tuples

$$\boldsymbol{u}_1 = (u_{10}, u_{11}, u_{12}, \ldots, u_{1r}), \boldsymbol{u}_2 = (u_{20}, u_{21}, u_{22}, \ldots, u_{2r}), \ldots, \boldsymbol{u}_s = (u_{s0}, u_{s1}, \ldots u_{sr}),$$

such that $b_j(v) \in N_{G_{n,\varepsilon}^+}(v)$, for $j \in [t_v]$ and $v \in [1, K]$, and $\boldsymbol{u}_a \in \mathcal{S}_r(G_{n,\varepsilon}^-)$, for $a \in [1, s]$.

Then expanding the product $T(K_{1,2}, G_{n,\varepsilon}^-)^s \prod_{v=1}^{K} T_{G_{n,\varepsilon}^+}(v)^{t_v}$ over the sum, the LHS of (2.9) can be bounded above by:

$$\sum_{\Gamma} \left| \mathbb{E}\left(\prod_{v=1}^{K} \prod_{j=1}^{t_v} \mathbf{1}\{X_v = X_{b_j(v)}\} \prod_{a=1}^{s} \mathbf{1}\{X_{=\boldsymbol{u}_a}\} \right) - \prod_{v=1}^{K} \left(\mathbb{E}\prod_{j=1}^{t_v} \mathbf{1}\{X_v = X_{b_j(v)}\} \mathbb{E}\prod_{a=1}^{s} \mathbf{1}\{X_{=\boldsymbol{u}_a}\} \right) \right|$$

$$= \sum_{\Gamma} \left| \beta\left(H_1 \bigcup H_2\right) - \beta(H_1)\beta(H_2) \right| \tag{2.13}$$

where $\beta(\cdot)$ is defined in (2.10) and

- H_1 is the simple labeled subgraph of $G_{n,\varepsilon}^+$ obtained by the union of the edges $(v, b_j(v))$ for $j \in [1, t_v]$ and $v \in [1, K]$.
- H_2 is the simple labeled subgraph of $G_{n,\varepsilon}^-$ obtained by the union of the r-stars formed by the collection of $(r+1)$-tuples $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_s\}$. More formally, $H_2 = (V(H_2), E(H_2))$, where

$$V(H_2) = \bigcup_{j=1}^{s} \boldsymbol{u}_j \quad \text{and} \quad E(H_2) = \bigcup_{j=1}^{s} \left\{(u_{j0}, u_{ja}) : 1 \le a \le r\right\}.$$

Note that if $V(H_1) \bigcap V(H_2) = \varnothing$, then $\beta\left(H_1 \bigcup H_2\right) = \beta(H_1)\beta(H_2)$, and so without loss of generality we may assume that the sum over Γ includes only terms for which $H_1 \bigcap H_2 \ne \varnothing$.

Definition 2.3. Let $\mathcal{H}_{m_1, m_2}$ denote the set of all unlabeled graphs $H = (V(H), E(H))$ which can be formed by the union of m_1 edges and m_2 copies of $K_{1,r}$.

Now, recalling that $\beta\left(H_1 \bigcup H_2\right) = \beta(H_1)\beta(H_2)$, if $V(H_1) \bigcap V(H_2) = \varnothing$, and $\beta\left(H_1 \bigcup H_2\right) \ge \beta(H_1)\beta(H_2)$ otherwise, the RHS of (2.13) can be bounded as follows:

$$\sum_{\Gamma} \left| \beta\left(H_1 \bigcup H_2\right) - \beta(H_1)\beta(H_2) \right| \le \sum_{\Gamma} \beta\left(H_1 \bigcup H_2\right)$$

$$= \sum_{m_1=1}^{\tilde{s}} \sum_{m_2=1}^{s} \sum_{H \in \mathcal{H}_{m_1, m_2}} \sum_{\Gamma : H_1 \bigcup H_2 \cong H} \beta\left(H_1 \bigcup H_2\right)$$

$$\lesssim \sum_{m_1=1}^{\tilde{s}} \sum_{m_2=1}^{s} \sum_{H \in \mathcal{H}_{m_1, m_2}} \beta(H) N(H, G_{n,\varepsilon}^+[K], G_{n,\varepsilon}^-), \tag{2.14}$$

where

- $G_{n,\varepsilon}^+[K]$ be the induced subgraph of $G_{n,\varepsilon}^+$ formed by the vertices labeled $\{1, 2, \ldots, K\}$ and its neighbors, that is, the 1-neighborhood of the K highest degree vertices in G_n; and
- $N(H, G_{n,\varepsilon}^+[K], G_{n,\varepsilon}^-)$ is the number of ways the graph H can be obtained as the union two graphs $H_1 \bigcup H_2$, such that H_1 is formed by the union of m_1 edges from $G_{n,\varepsilon}^+[K]$ and H_2 is formed by the union of m_2 copies of $K_{1,r}$ from $G_{n,\varepsilon}^-$, and $V(H_1) \bigcap V(H_2) \neq \emptyset$.
- $\tilde{s} := \sum_{v=1}^{K} t_v$.

Now, using $\beta(H) = \frac{1}{c_n^{|V(H)|-v(H)}}$ (by Lemma 2.5), and since the sum over m_1, m_2, H in (2.14) are all finite, to prove (2.9) it suffices to show that for every $H \in \mathcal{H}_{m_1,m_2}$,

$$\limsup_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{N(H, G_{n,\varepsilon}^+[K], G_{n,\varepsilon}^-)}{c_n^{V(H)-v(H)}} = 0. \tag{2.15}$$

To this end, fix $H \in \mathcal{H}_{m_1,m_2}$, and let H be isomorphic to $H_1 \bigcup H_2$ such that H_1 is formed by the union of m_1 edges from $G_{n,\varepsilon}^+[K]$ and H_2 is formed by the union of m_2 copies of $K_{1,r}$ from $G_{n,\varepsilon}^-$, and $V(H_1) \bigcap V(H_2) \neq \emptyset$. Such a pair (H_1, H_2) must exist, otherwise $N(H, G_{n,\varepsilon}^+[K], G_{n,\varepsilon}^-) = 0$. Let $C_1, C_2, \ldots, C_{v(H)}$ the connected components of H. Fix $1 \leq j \leq v(H)$ and consider the following three cases:

- $V(C_j)$ only intersects $V(H_1)$. Since $G_{n,\varepsilon}^+[K]$ is a bi-partite graph with bi-partition with $|E(G_{n,\varepsilon}^+[K])| \leq K\Delta(G_n) \lesssim_r Kc_n$ (using $\Delta(G_n) = O(c_n)$), this gives

$$N(C_j, G_{n,\varepsilon}^+[K]) \leq |E(G_{n,\varepsilon}^+[K])|^{|V(C_j)|-1} \lesssim_r (Kc_n)^{|V(C_j)|-1}.$$

- $V(C_j)$ only intersects $V(H_2)$. Then there exists $1 \leq h \leq m_2$ such that H_2 is spanned by h isomorphic copies of $K_{1,r}$. Thus, using the bound $N(K_{1,r}, G_n) = \Theta(c_n^r)$, gives

$$N(C_j, G_{n,\varepsilon}^-) \leq N(K_{1,r}, G_n)\Delta(G_n)^{|V(C_j)|-r+1} \lesssim_r c_n^{|V(C_j)|-1},$$

where the final step uses $\Delta(G_n) = O(c_n)$.

- $V(C_j)$ intersects both $V(H_1)$ and $V(H_2)$. Then, since C_j is connected, there exists vertices (u, v, w) such that $v \in V(H_1) \bigcap V(H_2)$, and (u, v) is an edge in $G_{n,\varepsilon}^+[K]$, and (v, w) is an edge $G_{n,\varepsilon}^-$. Thus, using the estimate $\Delta(G_n) = O(c_n)$,

$$N(C_j, G_{n,\varepsilon}^+[K], G_{n,\varepsilon}^-) \lesssim |E(G_{n,\varepsilon}^+[K])| \left(\max_{v \in V(G_{n,\varepsilon}^-)} d_v \right) \Delta(G_n)^{|V(C_j)|-3}$$

$$\lesssim_r K\varepsilon c_n^{|V(C_j)|-1}.$$

Taking a product over $1 \leq j \leq v(H)$ and, since $V(H_1) \bigcap V(H_2) \neq \emptyset$, gives

$$N(H, G_{n,\varepsilon}^+[K], G_{n,\varepsilon}^-) \lesssim_{r,m_1,m_2} \varepsilon K^{|V(H)|-v(H)} c_n^{|V(H)|-v(H)},$$

which implies (2.15), from which the desired conclusion follows. $\square$

2.3 | Contribution from $G_{n,\varepsilon}^+$

In this section we compute the asymptotic distribution of $T_+(K_{1,r}, G_{n,\varepsilon}^+)$ (recall (2.3)). This involves showing that the collection $\{T_{G_{n,\varepsilon}^+}(v) : 1 \leq v \leq K\}$ are asymptotically independent, by another moment comparison.

Lemma 2.6 *Assume (1.4) holds, and $\varepsilon > 0$ small enough. Then for all nonnegative integers $s_1, \ldots, s_K$,*

$$\lim_{n \to \infty} \left| \mathbb{E} \left(\prod_{v=1}^{K} T_{G_{n,\varepsilon}^+}(v)^{s_v} \right) - \prod_{v=1}^{K} \mathbb{E} T_{G_{n,\varepsilon}^+}(v)^{s_v} \right| = 0. \tag{2.16}$$

As a consequence, $T_+(K_{1,r}, G_{n,\varepsilon}^+) \overset{D}{\to} \sum_{v=1}^{\eta} \binom{T_v}{r}$, as $n \to \infty$, where $T_1, T_2, \ldots, T_\eta$ are independent $\mathrm{Pois}(\theta_1), \mathrm{Pois}(\theta_2), \ldots, \mathrm{Pois}(\theta_\eta)$, respectively. (Recall that $\eta = \eta(\varepsilon)$ is such that $\theta_{\eta+1} < \varepsilon < \theta_\eta$.)

Proof Expanding the moments, we have

$$\left| \mathbb{E} \prod_{v=1}^{K} T_{G_{n,\varepsilon}^+}(v)^{s_v} - \prod_{v=1}^{K} \mathbb{E} T_{G_{n,\varepsilon}^+}(v)^{s_v} \right| = \sum_{\Gamma} \left| \mathbb{E} \prod_{v=1}^{K} \prod_{j=1}^{s_v} \mathbf{1}\{X_v = X_{b_j(v)}\} - \prod_{v=1}^{K} \mathbb{E} \prod_{j=1}^{s_v} \mathbf{1}\{X_v = X_{b_j(v)}\} \right|$$

$$= \sum_{\Gamma} \left| \beta \left(\bigcup_{v=1}^{K} H(v) \right) - \prod_{v=1}^{K} \beta(H(v)) \right|$$

where

- Γ is the collection of all possible choices of $b_j(v) \in N_{G_{n,\varepsilon}}(v)$, for $j \in [s_v]$ and $v \in [K]$; and
- $H(v)$ denotes the simple graph formed by union of all the edges $(v, b_j(v))$, for $j \in [s_v]$. Note that $H(v)$ is isomorphic to a star graph, for every $v \in [K]$.

If $\bigcup_{v=1}^{K} H(v)$ is a forest, then the collection of random variables $\{\mathbf{1}\{X_v = X_{b_j(v)}, j \in [s_v], v \in [K]\}$ are mutually independent, and so, $\beta(\bigcup_{v=1}^{K} H(v)) = \prod_{v=1}^{K} \beta(H(v))$. Thus, without loss of generality, assume that $\bigcup_{v=1}^{K} H(v)$ is not a forest, that is, it contains a cycle. Then denoting $\mathcal{H}_m$ to be the set of unlabeled graphs with m vertices and $s := \sum_{v=1}^{K} s_v$, using Lemma 2.5 gives

$$\left| \mathbb{E} \prod_{v=1}^{K} T_{G_{n,\varepsilon}^+}(v)^{s_v} - \prod_{v=1}^{K} \mathbb{E} T_{G_{n,\varepsilon}^+}(v)^{s_v} \right| \lesssim \sum_{m=2}^{2s} \sum_{\substack{H \in \mathcal{H}_m \\ H \text{ contains a cycle}}} \sum_{\Gamma : \bigcup_{v=1}^{K} H(v) \simeq H} \beta \left(\bigcup_{v=1}^{K} H(v) \right)$$

$$= \sum_{m=2}^{2s} \sum_{\substack{H \in \mathcal{H}_m \\ H \text{ contains a cycle}}} N(H, G_{n,\varepsilon}^+[K]) \beta(H)$$

$$= \sum_{m=2}^{2s} \sum_{\substack{H \in \mathcal{H}_m \\ H \text{ contains a cycle}}} \frac{N(H, G_{n,\varepsilon}^+[K])}{c_n^{|V(H)|-v(H)}}. \tag{2.17}$$

Now, fix $H \in \mathcal{H}_m$ with connected components $H_1, H_2, \ldots, H_{v(H)}$, and assume without loss of generality that H_1 contains a cycle of length g. Also since H_1 is a subgraph of the bipartite graph $G_{n,\varepsilon}^+[K]$,

it follows that $g \geq 4$. Invoking [4, Lemma 2.3] gives,

$$N(H_1, G_{n,\varepsilon}^+[K]) \lesssim |E(G_{n,\varepsilon}^+[K])|^{|V(H_1)|-g/2} \lesssim (K\Delta(G_n))^{|V(H_1)|-g/2},$$

where the last inequality uses $|E(G_{n,\varepsilon}^+[K])| \leq K\Delta(G_n)$. Also, by [4, Lemma 2.3], for $j \geq 2$,

$$N(H_j, G_{n,\varepsilon}^+[K]) \lesssim |E(G_{n,\varepsilon}^+[K])|^{|V(H_j)|-1} \leq (K\Delta(G_n))^{|V(H_j)|-1}.$$

Taking a product over j and using $\Delta(G_n) = O(c_n)$, gives

$$N(H, G_{n,\varepsilon}^+[K]) \leq \prod_{j=1}^{\nu(H)} N(H_j, G_n) \lesssim K^{|V(H)|-\nu(H)} c_n^{|V(H)|-\nu(H)-(g/2-1)},$$

which implies $\limsup_{n\to\infty} \frac{N(H,G_{n,\varepsilon}^+[K])}{c_n^{|V(H)|-\nu(H)}} = 0$, as $g \geq 4$. Since the sum in (2.17) is finite (does not depend on n, ε), the conclusion in (2.16) follows.

Moreover, since $T_{G_{n,\varepsilon}^+}(v) \to \text{Pois}(\theta_v)$ in distribution and in moments, (2.16) implies that

$$\lim_{n\to\infty} \left| \mathbb{E}\left(\prod_{v=1}^{\eta} T_{G_{n,\varepsilon}^+}(v)^{s_v} \right) - \prod_{v=1}^{\eta} \mathbb{E} \, \text{Pois}(\theta_v)^{s_v} \right|.$$

This implies, as the Poisson distribution is uniquely determined by its moments,

$$(T_{G_{n,\varepsilon}^+}(1), T_{G_{n,\varepsilon}^+}(2), \ldots, T_{G_{n,\varepsilon}^+}(\eta)) \to (T_1, T_2, \ldots, T_\eta),$$

as $n \to \infty$, in distribution and in moments, where $T_1, T_2, \ldots, T_\eta$ are independent $\text{Pois}(\theta_1), \text{Pois}(\theta_2)$, $\ldots, \text{Pois}(\theta_\eta)$, respectively. Finally, recalling (2.8) and by the continuous mapping theorem $T_+(K_{1,r}, G_{n,\varepsilon}^+) = \sum_{v=1}^{\eta} \binom{T_{G_{n,\varepsilon}^+}(v)}{r} \to \sum_{v=1}^{\eta} \binom{T_v}{r}$ in distribution and in moments, as $n \to \infty$. ∎

2.4 | Contribution from $G_{n,\varepsilon}^-$

In this section we derive the limiting distribution of $T(K_{1,r}, G_{n,\varepsilon}^-)$, by invoking [5, Theorem 2.1], which gives conditions under which the number of monochromatic subgraphs (in particular monochromatic stars) converges to a linear combination of Poisson variables.

Lemma 2.7 *As $n \to \infty$ followed by $\varepsilon \to 0$,*

$$T(K_{1,r}, G_{n,\varepsilon}^-) \to \sum_{k=1}^{r+1} k Z_k,$$

in distribution and in moments, where $Z_1, Z_2, \ldots, Z_{r+1}$ are independent $\text{Pois}(\lambda_1 - \frac{1}{r!} \sum_{u=1}^{\infty} \theta_u^r)$, $\text{Pois}(\lambda_2), \ldots \text{Pois}(\lambda_{r+1})$, respectively.

Proof of Lemma 2.7

We will prove this result by invoking [5, Theorem 2.1] which we restate in Appendix A (Theorem A.1) adapted to the case of r-stars, for the sake of completeness.

To begin with, let F be a graph formed by the *join* of two isomorphic copies of $K_{1,r}$, such that $|V(F)| > r + 1$ (refer to Definition A.1 for a formal definition). This ensures that F is a t-join of $K_{1,r}$, for some $t \in [2, r]$. Moreover, F is connected by definition, and

$$N(F, G_{n,\varepsilon}^-) \lesssim N(K_{1,r}, G_{n,\varepsilon}^-) \cdot \Delta(G_{n,\varepsilon}^-)^{|V(F)|-r-1} \leq N(K_{1,r}, G_n) \cdot (\varepsilon c_n)^{|V(F)|-r-1}$$

$$\lesssim \varepsilon^{|V(F)|-r-1} c_n^{|V(F)|-1}.$$

Therefore,

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{c_n^{|V(F)|-1}} N(F, G_{n,\varepsilon}^-) = 0, \tag{2.18}$$

when $|V(F)| > r + 1$. This verifies condition (A2).

To invoke Theorem A.1, it remains to establish (A1). To this end, consider super-graphs $F \supseteq K_{1,r}$ with $|V(F)| = r + 1$. Recalling $\mathscr{C}_{r,k} := \{F \supseteq K_{1,r} : |V(F)| = r + 1 \text{ and } N(K_{1,r}, F) = k\}$, we have the following lemma.

Lemma 2.8 *For any $F \in \mathscr{C}_{r,k}$, with $k \in [2, r+1]$, $N_{\text{ind}}(F, G_{n,\varepsilon}) = N_{\text{ind}}(F, G_{n,\varepsilon}^-) + o(c_n^r)$, as $n \to \infty$ followed by $\varepsilon \to 0$.*

Proof Let $k \in [2, r+1]$ and suppose $F \in \mathscr{C}_{r,k}$ is an induced subgraph of $G_{n,\varepsilon}$, such that $V(F)$ is not completely contained in $V(G_{n,\varepsilon}^-)$. Then, since F has at least two vertices of degree r and any two degree r vertices must be neighbors, the vertices of F can be spanned by a r-star whose central vertex is in $N_{G_{n,\varepsilon}}(V_\varepsilon(G_n))$. (Note that by construction there are no edges between the vertices of $V_\varepsilon(G_n)$ in the graph $G_{n,\varepsilon}$, and so $N_{G_{n,\varepsilon}}(V_\varepsilon(G_n))$ does not intersect $V_\varepsilon(G_n)$.) Therefore, the difference $N_{\text{ind}}(F, G_{n,\varepsilon}) - N_{\text{ind}}(F, G_{n,\varepsilon}^-)$ is bounded above by (up to constants depending only on r)

$$\sum_{v \notin V_\varepsilon(G_n)} \sum_{u_1 \in V_\varepsilon(G_n)} \sum_{\boldsymbol{u} \in \binom{V(G_n)}{r-1}} a_{vu_1}(G_n) a_v(\boldsymbol{u}, G_n) \leq (\varepsilon c_n)^{r-1} \sum_{v \notin V_\varepsilon(G_n)} \sum_{u_1 \in V_\varepsilon(G_n)} a_{vu_1}(G_n) \qquad \text{(by (2.5))}$$

$$\leq (\varepsilon c_n)^{r-1} \sum_{u_1 \in V_\varepsilon(G_n)} d_{u_1},$$

which is $o(c_n^r)$ (as $n \to \infty$ followed by $\varepsilon \to 0$), by (2.6), and the argument following it. ∎

Using the above lemma and $N_{\text{ind}}(F, G_n) = N_{\text{ind}}(F, G_{n,\varepsilon}) + o(c_n^r)$ (by Lemma 2.2), it follows that, for $k \in [2, r+1]$,

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\sum_{F \in \mathscr{C}_{r,k}} N_{\text{ind}}(F, G_{n,\varepsilon}^-)}{c_n^r} = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\sum_{F \in \mathscr{C}_{r,k}} N_{\text{ind}}(F, G_n)}{c_n^r} = \lambda_k, \tag{2.19}$$

where the last equality uses (1.5). This establishes (A1) for $k \in [2, r+1]$.

It remains to consider the case $k = 1$. To begin with, observe that for any graph G,

$$N(K_{1,r}, G) = \sum_{k=1}^{r+1} \sum_{F \in \mathscr{C}_{r,k}} k N_{\text{ind}}(F, G). \tag{2.20}$$

Moreover, using Lemmas 2.2 and 2.3 gives

$$N(K_{1,r}, G_{n,\varepsilon}^-) = N(K_{1,r}, G_n) - \sum_{v=1}^{\eta} \binom{d_v}{r} + o(c_n^r).$$

Now, using this and (2.20) with $G = G_{n,\varepsilon}^-$ gives

$$\frac{\sum_{F \in \mathscr{C}_{r,1}} N_{\text{ind}}(F, G_{n,\varepsilon}^-)}{c_n^r} = \frac{N(K_{1,r}, G_n)}{c_n^r} - \frac{1}{c_n^r} \sum_{v=1}^{\eta} \binom{d_v}{r} - \sum_{k=2}^{r+1} k \sum_{F \in \mathscr{C}_{r,k}} \frac{N_{\text{ind}}(F, G_{n,\varepsilon}^-)}{c_n^r} + o(1)$$

$$\rightarrow \sum_{k=1}^{r+1} k\lambda_k - \sum_{u=1}^{\infty} \frac{\theta_u^r}{r!} - \sum_{k=2}^{r+1} k\lambda_k \qquad \text{(using (2.20) with } G = G_n \text{ and (1.5))}$$

$$= \lambda_1 - \sum_{u=1}^{\infty} \frac{\theta_u^r}{r!}, \tag{2.21}$$

as $n \rightarrow \infty$ followed by $\varepsilon \rightarrow 0$. This establishes (A1) for $k = 1$.

Finally, combining (2.18), (2.19), and (2.21) and using Theorem A.1, we have $T(K_{1,r}, G_{n,\varepsilon}^-) \xrightarrow{D} \sum_{k=1}^{r+1} kZ_k$, where $Z_1, Z_2, \ldots, Z_{r+1}$ are as in the statement of the lemma. The convergence in moments is a consequence of uniform integrability as $\mathbb{E}(T(K_{1,r}, G_{n,\varepsilon}^-)) \leq \mathbb{E}T(K_{1,r}, G_n)^r = O_r(1)$ for every fixed integer $r \geq 1$ [4, Theorem 1.2].

2.5 | Completing the proof of Theorem 1.2

We begin with the proof of (a). Note that by Lemma 2.3 it suffices to find the limiting distribution of

$$\sum_{v=1}^{\eta(\varepsilon)} \binom{T_{G_{n,\varepsilon}^+}(v)}{r} + T(K_{1,r}, G_{n,\varepsilon}^-), \tag{2.22}$$

under the double limit as $n \rightarrow \infty$ followed by $\varepsilon \rightarrow 0$. Fix an integer $K \geq 1$ and write the above random variable as

$$\sum_{v=1}^{K} \binom{T_{G_{n,\varepsilon}^+}(v)}{r} + \sum_{v=K+1}^{\eta(\varepsilon)} \binom{T_{G_{n,\varepsilon}^+}(v)}{r} + T(K_{1,r}, G_{n,\varepsilon}^-).$$

Under the double limit the random vector

$$\left(T_{G_{n,\varepsilon}^+}(1), \ldots, T_{G_{n,\varepsilon}^+}(K), T(K_{1,r}, G_{n,\varepsilon}^-) \right) \xrightarrow{D} \left(T_1, \ldots, T_K, \sum_{k=1}^{r+1} kZ_k \right),$$

by invoking Lemmas 2.4, 2.6, and 2.7. By continuous mapping theorem this gives

$$\sum_{v=1}^{K} \binom{T_{G_{n,\varepsilon}^+}(v)}{r} + T(K_{1,r}, G_{n,\varepsilon}^-) \xrightarrow{D} \sum_{v=1}^{K} \binom{T_v}{r} + \sum_{k=1}^{r+1} kZ_k, \tag{2.23}$$

the RHS of which on letting $K \to \infty$ converges in distribution to $\sum_{v=1}^{\infty} \binom{T_v}{r} + \sum_{k=1}^{r+1} k Z_k$. It thus suffices to show that

$$\lim_{K \to \infty} \lim_{\varepsilon \to 0} \lim_{n \to \infty} \sum_{v=K+1}^{\eta} \mathbb{E} \binom{T_{G_{n,\varepsilon}^+}(v)}{r} = 0. \tag{2.24}$$

The LHS above is bounded above by $\sum_{v=K+1}^{\eta} \frac{1}{r!} \frac{d_v^r}{c_n^r}$, which on letting $n \to \infty$ followed by $\varepsilon \to 0$ gives $\frac{1}{r!} \sum_{v=K+1}^{\infty} \theta_v^r$. This converges to 0 as $K \to \infty$, as $\sum_{v=1}^{\infty} \theta_v^r < \infty$, as noted in the proof of Lemma 2.3.[5] Combining (2.22), (2.23), and (2.24) gives

$$\sum_{v=1}^{\eta(\varepsilon)} \binom{T_{G_{n,\varepsilon}^+}(v)}{r} + T(K_{1,r}, G_{n,\varepsilon}^-) \overset{D}{\to} \sum_{v=1}^{\infty} \binom{T_v}{r} + \sum_{k=1}^{r+1} k Z_k, \tag{2.25}$$

which proves the distributional convergence in (1.7). (Note that the proof above works for the case $\|\theta\|_\infty > 0$. If $\|\theta\|_\infty = 0$, the proof follows on noting that $T(K_{1,r}, G_n) = T(K_{1,r}, G_{n,\varepsilon}^-) + o_P(1)$, by Lemma 2.1, and then invoking Lemma 2.7.)

Finally, to complete the proof of Theorem 1.2(a) we need to show convergence in moments, which is a consequence of uniform integrability as all moments of $T(K_{1,r}, G_n)$ are bounded: that is, $\mathbb{E} T(K_{1,r}, G_n)^r = O_r(1)$ for every fixed integer $r \geq 1$ (this follows from the proof of [4, Theorem 1.2]).

Next, we prove the converse (b). By invoking Proposition 1.1 we can assume, without loss of generality, that $N(K_{1,r}, G_n) = O(c_n^r)$. This in turn implies that for every graph F on $r+1$ vertices which is a super graph of $K_{1,r}$ we have $N_{\text{ind}}(F, G_n) = O(c_n^r)$. Thus by passing to a subsequence, assume that $N_{\text{ind}}(F, G_n)/c_n^r$ converges for every F which is a super graph of $K_{1,r}$. This implies existence of the limits in (1.5). Finally, using (2.2) we have $\max_{v \in V(G_n)} d_v = O(c_n)$, and so the infinite tuple $\{d_v/c_n\}_{v \geq 1}$ is an element of $[0, K]^{\mathbb{N}}$ for some K fixed. Since $[0, K]^{\mathbb{N}}$ is compact in product topology, there is a further subsequence along which d_v/c_n converges for every $v \geq 1$ simultaneously. Thus, moving to a subsequence, we can assume that d_v/c_n converges to θ_v for every v. Invoking the sufficiency part of the theorem gives that $T(K_{1,r}, G_n)$ converges in distribution to a random variable of the desired form, completing the proof.

2.6 | Proof of Corollary 1.3

The proof of $(a) \Rightarrow (b)$ is immediate from Theorem 1.2, so it suffices to prove $(b) \Rightarrow (a)$. To this end, note that $T(K_{1,r}, G_n) \overset{D}{\to} \sum_{k=1}^{r+1} k Z_k$ implies that (1.4) holds (Proposition 1.1). Thus, by a similar argument which was used to prove the converse of Theorem 1.2, it follows that along a subsequence the limits $\lim_{n \to \infty} \frac{1}{c_n^r} N_{\text{ind}}(F, G_n)$ exist for all super graphs F of $K_{1,r}$ on $r+1$ vertices, and so, for $k \in [1, r+1]$,

$$\lambda'_k := \lim_{n \to \infty} \sum_{F \in \mathcal{C}_{r,k}} \frac{N_{\text{ind}}(F, G_n)}{c_n^r}$$

is well defined. Then, as before, by passing to another subsequence the limits $\theta'_v := \lim_{n \to \infty} \frac{d_v}{c_n}$ exist for every $v \geq 1$, and by the if part of Theorem 1.2 along this subsequence,

$$T(K_{1,r}, G_n) \overset{d}{\to} \sum_{v=1}^{\infty} \binom{T'_v}{r} + \sum_{k=1}^{r+1} k Z'_k,$$

---

[5]Note that if $\eta_0 = \lim_{\varepsilon \to 0} \eta(\varepsilon) < \infty$, then the term $\sum_{v=M+1}^{\eta} \binom{T_{G_{n,\varepsilon}^+}(v)}{r} + T(K_{1,r}, G_{n,\varepsilon}^-)$ vanishes for $M = \eta_0$, thus simplifying the proof.

where $\{T_v'\}_{v \geq 1}$ and $\{Z_k'\}_{1 \leq k \leq r+1}$ are mutually independent, and $T_1', T_2', \ldots,$ are independent $\mathrm{Pois}(\theta_1')$, $\mathrm{Pois}(\theta_2')$, ..., respectively, and $Z_1', Z_2', \ldots, Z_{r+1}'$ are independent $\mathrm{Pois}(\lambda_1' - \frac{1}{r!} \sum_{u=1}^{\infty} (\theta_u')^r)$, $\mathrm{Pois}(\lambda_2')$, $\ldots, \mathrm{Pois}(\lambda_{r+1}')$, respectively.

However, since $T(K_{1,r}, G_n)$ converges in distribution to $\sum_{k=1}^{r+1} k Z_k$ which has finite exponential moment everywhere, it follows that $\theta_v' = 0$ for all $v \geq 1$, and consequently, the maximum degree $\Delta(G_n) = o(c_n)$. This also gives

$$\sum_{k=1}^{r+1} k Z_k \stackrel{D}{=} \sum_{k=1}^{r+1} k Z_k',$$

and so the corresponding probability generating functions must match, that is,

$$\prod_{k=1}^{r+1} e^{\lambda_k (s^k - 1)} = \prod_{k=1}^{r+1} e^{\lambda_k' (s^k - 1)}, \quad \text{for all } s \in (0, 1).$$

This implies, $\sum_{k=1}^{r+1} \lambda_k (s^k - 1) = \sum_{k=1}^{r+1} \lambda_k' (s^k - 1)$, for all $s \in (0, 1)$, and so the corresponding coefficients must be equal, giving $\lambda_k = \lambda_k'$. Therefore, every sub sequential limit of $\sum_{F \in \mathcal{C}_{r,k}} \frac{N_{\mathrm{ind}}(K_{1,r}, G_n)}{c_n^r}$ equal λ_k, for $k \in [1, r+1]$, hence, (1.5) holds.

3 | EXAMPLES

In this section we apply Theorem 1.2 to different deterministic and random graph models, and determine the specific nature of the limiting distribution.

Example 1. (Disjoint union of stars) The proof of Theorem 1.2 shows that the quadratic term in the limiting distribution of $T(K_{1,r}, G_n)$ appears due to the r-stars incident on vertices with degree $\Theta(c_n)$. This can be seen when G_n is a disjoint union of star graphs.

- To begin with suppose $G_n = K_{1,n}$ is the n-star. Then $N(K_{1,r}, K_{1,n}) = \binom{n}{r}$, and if we color $K_{1,n}$ with c_n colors such that $n/c_n \to 1$, then $\mathbb{E}(T(K_{1,r}, G_n)) = \frac{1}{r!}$. Note that the maximum degree $d_{(1)} = n$, which implies $\theta_1 = 1$. Moreover, $d_{(2)} = 1$, which implies $\theta_v = 0$, for all $v \geq 2$. Therefore, by Theorem 1.2,

$$T(K_{1,r}, G_n) \stackrel{D}{\to} \binom{T_1}{r},$$

where $T_1 \sim \mathrm{Pois}(1)$. (Note that the graph $G_{n,\varepsilon}^-$ is empty in this case.)

- Next, consider G_n to be the disjoint union of the following stars: $K_{1,\lfloor na_1 \rfloor}, K_{1,\lfloor na_2 \rfloor}, \ldots, K_{1,\lfloor na_n \rfloor}$, such that $\sum_{s=1}^{\infty} a_s^r < \infty$. In this case, $N(K_{1,r}, G_n) = \sum_{s=1}^{n} \binom{\lfloor na_s \rfloor}{r} \sim \frac{n^r}{r!} \sum_{s=1}^{n} a_s^r$. If G_n is colored with c_n colors such that $n/c_n \to 1$, then $\mathbb{E}(T(K_{1,r}, G_n)) \to \frac{1}{r!} \sum_{s=1}^{\infty} a_s^r$. Also, $d_{(v)} = \lfloor na_v \rfloor$, which implies $\theta_v = a_v$, for $v \geq 1$. This implies, by Theorem 1.2,

$$T(K_{1,r}, G_n) \stackrel{D}{\to} \sum_{s=1}^{\infty} \binom{T_s}{r},$$

where $T_s \sim \mathrm{Pois}(a_s)$ and $T_1, T_2, \ldots$ are independent. Here, the linear terms linear in Poisson do not contribute, as $G_{n,\varepsilon}^-$ is empty, and $\mathbb{E}T(K_{1,r}, G_n) \sim \frac{1}{r!} \sum_{v=1}^{\infty} \theta_v^r$.

- Finally, consider G_n to be the disjoint union of the following stars:

$$K_{1,\lfloor na_1+n^{\frac{r-1}{r}}\rfloor}, K_{1,\lfloor na_2+n^{\frac{r-1}{r}}\rfloor}, \ldots, K_{1,\lfloor na_n+n^{\frac{r-1}{r}}\rfloor}.$$

In this case,

$$N(K_{1,r}, G_n) = \sum_{s=1}^n \binom{\lfloor na_s + n^{\frac{r-1}{r}}\rfloor}{r} \sim \frac{n^r}{r!} + \frac{n^r}{r!}\sum_{s=1}^n a_s^r,$$

since $\sum_{s=1}^n a_s^k = o(n^{1-\frac{k}{r}})$, for $1 \le k < r$ (see Observation 3.1). If G_n is colored with c_n colors such that $n/c_n \to 1$, then $\mathbb{E}(T(K_{1,r}, G_n)) \to \frac{1}{r!}\left(1 + \sum_{s=1}^\infty a_s^r\right)$. Also, $d_{(v)} = \lfloor na_v + n^{\frac{r-1}{r}}\rfloor$, which implies $\theta_v = a_v$, for $v \ge 1$, and so Theorem 1.2 gives

$$T(K_{1,r}, G_n) \overset{D}{\to} \sum_{s=1}^\infty \binom{T_s}{r} + Z,$$

where $T_s \sim \text{Pois}(a_s)$ and $T_1, T_2, \ldots$ are independent, and $Z \sim \text{Pois}(\frac{1}{r!})$ independent of $\{T_s\}_{s\ge1}$.

Observation 3.1   *If $\{a_s\}_{s\ge1}$ is a sequence of nonnegative real numbers such that $\sum_{s=1}^\infty a_s^r < \infty$ then $\sum_{s=1}^n a_s^k = o(n^{1-\frac{k}{r}})$, for $1 \le k < r$.*

Proof   Fixing $\varepsilon > 0$ and a positive integer $N \ge 1$ we get

$$\sum_{s=1}^n a_s^k = \sum_{s=1}^N a_s^k + \sum_{s=N+1}^n a_s^k \mathbf{1}\{a_s \le \varepsilon n^{-\frac{1}{r}}\} + \sum_{s=N+1}^n a_s^k \mathbf{1}\{a_s > \varepsilon n^{-\frac{1}{r}}\}$$

$$\le \sum_{s=1}^N a_s^k + \varepsilon^k n^{1-\frac{k}{r}} + \frac{n^{1-\frac{r}{k}}}{\varepsilon^{r-k}}\sum_{s=N+1}^\infty a_s^r.$$

On dividing by $n^{1-\frac{k}{r}}$ and letting $n \to \infty$, the first term goes to 0 as it is a finite sum, and, therefore,

$$\limsup_{n\to\infty} \frac{\sum_{s=1}^k a_s^r}{n^{1-\frac{k}{r}}} \le \varepsilon^k + \frac{1}{\varepsilon^{r-k}}\sum_{s=N+1}^\infty a_s^r.$$

The desired conclusion now follows on letting $N \to \infty$ followed by $\varepsilon \to 0$, on noting that $\sum_{s=1}^\infty a_s^r < \infty$. ∎

Next, we see examples where there are no vertices of high degree, in which case, the quadratic term vanishes (Corollary 1.3).

Example 2. (Regular graphs) Let G_n be a d-regular graph. In this case, $N(K_{1,r}, G_n) = n\binom{d}{r}$. Consider uniformly coloring the graph with c_n colors such that $\frac{1}{c_n^r}n\binom{d}{r} \to \lambda$. In this case, $\Delta(G_n) = \max_{v\in V(G_n)} d_v = d = o(c_n)$. Therefore, by Corollary 1.3, $T(K_{1,r}, G_n) \overset{D}{\to} \sum_{k=1}^{r+1} kZ_k$, where $Z_1, Z_2, \ldots, Z_{r+1}$ are independent $\text{Pois}(\lambda_1), \text{Pois}(\lambda_2), \ldots, \text{Pois}(\lambda_{r+1})$ (recall (1.5)). (Note that $\sum_{k=1}^{r+1} k\lambda_k = \lambda$.) The limit simplifies in special cases:
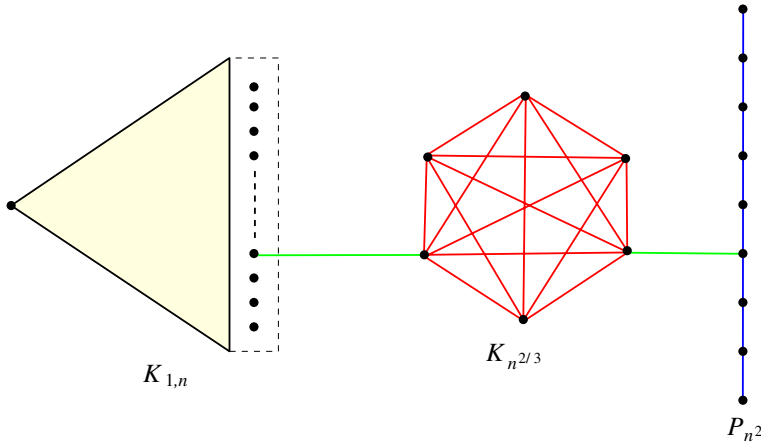
FIGURE 2 Illustration for Example 3 [Color figure can be viewed at wileyonlinelibrary.com]

- $G_n = K_{n,n}$, the regular bipartite graph. Since, bipartite graphs are triangle-free, $N_{\text{ind}}(F, G_n) = 0$, for any super-graph F of $K_{1,r}$ with $|V(F)| = r + 1$. This implies $\lambda_k = 0$, for $2 \le k \le r + 1$, and $\lambda_1 = \lambda$, and $T(K_{1,r}, K_{n,n}) \xrightarrow{D} \text{Pois}(\lambda)$.
- $G_n = K_n$, the complete graph on n vertices. In this case, any induced graph on $r + 1$ vertices is isomorphic to K_{r+1}. This implies $\lambda_k = 0$, for $1 \le k \le r$ and $\lambda_{r+1} = \frac{\lambda}{r+1}$, and $T(K_{1,r}, K_n) \xrightarrow{D} (r+1)Z_{r+1}$, where $Z_{r+1} \sim \text{Pois}(\frac{\lambda}{r+1})$.

Note that in all the above examples, the limiting distribution either involves only the quadratic part or only the linear part. It is easy to construct examples where both the components show up by taking disjoint unions (or connecting them with a few edges) of the graphs in the above examples, as shown below:

Example 3. Let G_n be the graph in Figure 2. Note that it has three parts, a $K_{1,n}$, where one of the leaves is connected by a single edge to a $K_{n^{2/3}}$, which is connected by a single edge to a path P_{n^2}. Consider coloring this graph by c_n colors such that $c_n/n \to \kappa$. This implies

$$\mathbb{E}(T(K_{1,2}, G_n)) = \frac{1}{c_n^2} N(K_{1,2}, G_n) \sim \frac{\binom{n}{2} + 3\binom{\lceil n^{2/3} \rceil}{3} + n^2}{c_n^2} \to 2\kappa^2.$$

Next, note that $\Delta(G_n) = n$, which corresponds to the central vertex of the $K_{1,n}$. Therefore, $\theta_1 = \kappa$. For every other vertex the degree is $o(n)$, which implies $\theta_v = 0$, for all $v \ge 2$. Finally, since $N(K_3, G_n) = \binom{\lceil n^{2/3} \rceil}{3}$, $\nu := \lim_{n \to \infty} \frac{1}{c_n^2} N(K_3, G_n) = \frac{\kappa^2}{6}$. Therefore, by Theorem 1.2

$$T(K_{1,2}, G_n) \xrightarrow{D} \binom{T_1}{2} + 3Z_3 + Z_1,$$

where $T_1 \sim \text{Pois}(\kappa)$, $Z_3 \sim \text{Pois}(\frac{\kappa^2}{6})$, and $Z_1 \sim \text{Pois}(\frac{\kappa^2}{2})$.

Remark 3.1. (Extension to random graphs) By a simple conditioning argument, Theorem 1.2 can be extended to random graphs by conditioning on the graph, under the assumption that the graph

and its coloring are jointly independent (see [5, Lemma 4.1]). In this case, whenever the limits in (1.4) and (1.6) exist in probability, the limit (1.7) holds. For example, when $G_n \sim G(n, p(n))$ is the Erdős-Rényi random graph, then the limiting distribution of $T(K_{1,r}, G_n)$ (when c_n is chosen such that $\frac{1}{c_n^r} \mathbb{E}(N(K_{1,r}, G_n)) \to \lambda$) can be easily derived using Theorem 1.2. In this case, depending on whether (a) $n^{\frac{r+1}{r}} p(n) \to O(1)$, (b) $p(n) \to 0$, $n^{\frac{r+1}{r}} p(n) \to \infty$, or (c) $p(n) = p \in (0, 1)$ is fixed, $T(K_{1,r}, G_n)$ converges to (a) zero in probability, or (b) Pois(λ), or (c) a linear combination of independent Poisson variables (see [5, Theorem 1.3] for details).

4 | CONCLUSION AND OPEN PROBLEMS

This paper studies the limiting distribution of the number of monochromatic r-stars in a uniformly random coloring of a growing graph sequence. We provide a complete characterization of the limiting distribution of $T(K_{1,r}, G_n)$, in the regime where $\mathbb{E}(T(K_{1,r}, G_n)) = \Theta(1)$.

It remains open to understand the limiting distribution of $T(K_{1,r}, G_n)$ when $\mathbb{E}(T(K_{1,r}, G_n)) = \frac{1}{c_n^r} N(K_{1,r}, G_n)$ grows to infinity. For the case of monochromatic edges, [4, Theorem 1.2] showed that $T(K_2, G_n)$ (centered by the mean and scaled by the standard deviation) converges to $N(0, 1)$, whenever $\mathbb{E}(T(K_2, G_n)) = \frac{1}{c_n} |E(G_n)| \to \infty$ such that $c_n \to \infty$. Error rates for the above CLT were obtained by Fang [14]. It is natural to wonder whether this universality phenomenon extends to monochromatic r-stars, and more generally, to any fixed connected graph H.

On the other hand, when $\mathbb{E}(T(K_2, G_n)) \to \infty$ such that the number of colors $c_n = c$ is fixed, then $T(K_2, G_n)$ (after appropriate centering and scaling) is asymptotically normal if and only if its fourth moment converges to 3 [4, Theorem 1.3]. It would be interesting to explore whether this fourth-moment phenomenon extends to monochromatic r-stars.

REFERENCES

1. A.D. Barbour, L. Holst, and S. Janson, *Poisson approximations*, Oxford University Press, Oxford, 1992.
2. A. Basak and S. Mukherjee, *University of the mean-field for the Potts model*, Probab. Theory Related Fields. **168** (2017), no. 3–4, 557–600.
3. S. Bhamidi, J.M. Steele, and T. Zaman, *Twitter event networks and the superstar model*, Ann. Appl. Probab. **25** (2015), no. 5, 2462–2502.
4. B.B. Bhattacharya, P. Diaconis, and S. Mukherjee, *Universal Poisson and normal limit theorems in graph coloring problems with connections to extremal combinatorics*, Ann. Appl. Probab. **27** (2017a), no. 1, 337–394.
5. B. B. Bhattacharya, S. Mukherjee, and S. Mukherjee, Birthday paradox, monochromatic subgraphs, and the second moment phenomenon. arXiv: 1711.01465, 2017b.
6. B. B. Bhattacharya and S. Mukherjee, Monochromatic subgraphs in randomly colored graphons. arXiv: 1707.05889, 2017.
7. A. Cerquetti and S. Fortini, *A Poisson approximation for colored graphs under exchangeability*, Sankhya. **68** (2006), no. 2, 183–197.

8. S. Chatterjee and P. Diaconis, *Estimating and understanding exponential random graph models*, Ann. Stat. **41** (2013), no. 5, 2428–2461.

9. S. Chatterjee, P. Diaconis, and E. Meckes, *Exchangeable pairs and Poisson approximation*, Electron. Encyclopedia Probab. (2004).

10. A. DasGupta, *The matching, birthday and the strong birthday problem: a contemporary review*, J. Statist. Plann. Inference. **130** (2005), 377–389.

11. P. Diaconis and S. Holmes, *A Bayesian peek into Feller I*, Sankhya, Series A. **64** (2002), no. 3, 820–841.

12. P. Diaconis and F. Mosteller, *Methods for studying coincidences*, J. Amer. Statist. Assoc. **84** (1989), no. 408, 853–861.

13. F.M. Dong, K.M. Koh, and K.L. Teo, *Chromatic polynomials and chromaticity of graphs*, World Scientific Publishing Company, Singapore, 2005.

14. 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.

15. J.H. Friedman and L.C. Rafsky, *Multivariate generalizations of the Wolfowitz and Smirnov two-sample tests*, Ann. Stat. **7** (1979), 697–717.

16. S.D. Galbraith and M. Holmes, *A non-uniform birthday problem with applications to discrete logarithms*, Discrete Appl. Math. **160** (2012), no. 10–11, 1547–1560.

17. T. Jensen and B. Toft. Unsolved graph colouring problems. In L. Beineke & R. Wilson (Eds.), *Topics in Chromatic Graph Theory (Encyclopedia of Mathematics and its Applications*, pp. 327–357). Cambridge: Cambridge University Press. (2015).

18. S. Mukherjee, Phase transition in the two star exponential random graph model. arXiv: 310.4164, 2013.

19. J. Park and M.E.J. Newman, *Solution of the 2-star model of a network*, Phys. Rev. E. **70** (2004), no. 6, 066146.

20. R.P. Stanley, *A symmetric function generalization of the chromatic polynomial of a graph*, Adv. Math. **111** (1995), no. 1, 166–194.

APPENDIX: MISSING INGREDIENTS FOR THE PROOF OF LEMMA 2.7

Here, we recall [4, Theorem 2.1], a limit theorem for general monochromatic subgraphs, which is used in the proof of Lemma 2.7. We begin with a few definitions. For a finite set S and a positive integer N, denote by S_N the set of all r-tuples $s = (s_1, \ldots, s_r) \in S^N$ with distinct entries.[6] Next, we need to define the notion of join of graphs.

Definition A.1. Let $H = (V(H), E(H))$ be a fixed graph and $t \in [1, |V(H)|]$. Let H' be an isomorphic copy of H, with $V(H) = \{1, 2, \ldots, |V(H)|\}$ and $V(H') = \{1', 2', \ldots, |V(H)|'\}$, where $z' \in V(H')$ is the image of $z \in V(H)$. For two ordered index sets $J_1 = (j_{11}, j_{12}, \ldots, j_{1t}) \in [|V(H)|]_t$ and $J_2 = (j_{21}, j_{22}, \ldots, j_{2t}) \in [|V(H)|]_t$, denote by $H_t(J_1, J_2)$ the simple graph obtained by the union of H and H', when the vertex $j_{1a} \in V(H)$ is identified with the vertex $j'_{2a} \in V(H')$, for $a \in [t]$. More precisely,

$$H_t(J_1, J_2) = \left(V(H) \bigcup \gamma(V(H')), E(H) \bigcup \gamma(E(H')) \right),$$

where

---

[6]For a set S, the set S^N denotes the N-fold Cartesian product $S \times S \times \cdots \times S$.
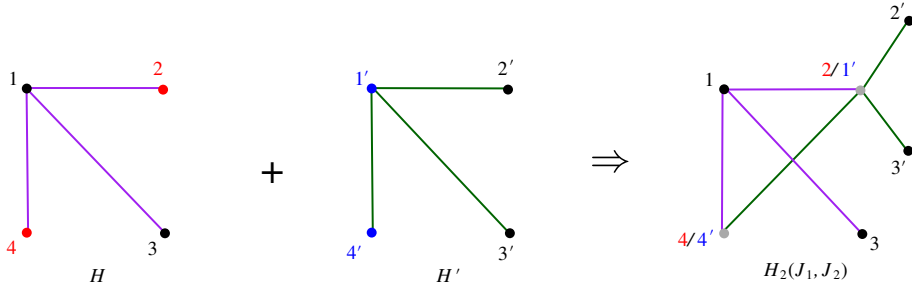
FIGURE 3 2-Join of $H = K_{1,3}$ (the 3-star) with pivots $J_1 = (2, 4)$ and $J_2 = (1, 4)$ [Color figure can be viewed at wileyonlinelibrary.com]

- $\gamma(V(H')) = \{\gamma(v') : v' \in V(H')\}$, where γ is a relabeling of the vertices of $V(H')$ such that $\gamma(j'_{2a}) = j_{1a}$, for $a \in [t]$, and $\gamma(v') = v'$ otherwise.
- This induces a relabeling of the edges $\gamma(E(H')) = \{\gamma((u', v')) : (u', v') \in E(H')\}$, where $\gamma((u', v')) = (\gamma(u'), \gamma(v'))$, for $(u', v') \in E(H')$.

The graph $H_t(J_1, J_2)$ will be referred to as the *t-join of H with pivots at J_1 and J_2* (see Figure 3). Denote by $\mathscr{J}_t(H) := \{H_t(J_1, J_2) : J_1, J_2 \in [|V(H)|]_t\}$ the collection of all graphs (up to isomorphism) which can be obtained as the *t-join of H*. Finally, a graph F is said to be a *join of two isomorphic copies of H*, if $F \in \mathscr{J}_t(H)$, for some $t \in [1, |V(H)|]$.

Now, we can restate [4, Theorem 2.1] specialized to the case of the *r-star*. Note that this result captures the asymptotic regime where $T(K_{1,r}, G_n)$ is asymptotically 'linear'. We apply this result in the proof of Lemma 2.7 for the sequence of graphs $G_{n,\varepsilon}$.

Theorem A.1 [4, Theorem 2.1] *Suppose W_n is a sequence of graphs colored uniformly with c_n colors, such that the following hold:*

- *For every $k \in [1, r + 1]$, there exists $\lambda_k \geq 0$ such that*

$$\lim_{n \to \infty} \frac{\sum_{F \in \mathscr{C}_{r,k}} N_{\text{ind}}(F, W_n)}{c_n^r} = \lambda_k, \tag{A1}$$

where $\mathscr{C}_{r,k} := \{F \supseteq K_{1,r} : |V(F)| = r + 1 \text{ and } N(K_{1,r}, F) = k\}$.
- *For $t \in [2, r]$ and every $F \in \mathscr{J}_t(K_{1,r})$, as $n \to \infty$,*

$$N(F, W_n) = o(c_n^{|V(F)|-1}). \tag{A2}$$

Then

$$T(K_{1,r}, W_n) \xrightarrow{D} \sum_{k=1}^{r+1} kZ_k,$$

where $Z_k \sim \text{Pois}(\lambda_k)$ and the collection $\{Z_k : 1 \leq k \leq r + 1\}$ is independent.