Variance of the subgraph count for sparse Erdős-Rényi graphs

Robert B. Ellis [a,1,*], James P. Ferry [b,1]

[a] *Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616*

[b] *Metron, Inc., 11911 Freedom Drive Suite 800 Reston, VA 20190-5602*

Abstract

We develop formulas for the variance of the number of copies of a small subgraph H in the Erdős-Rényi random graph. The central technique employs a graph overlay polynomial encoding subgraph symmetries, which is of independent interest, that counts the number of copies $\widetilde{H} \cong H$ overlapping H. In the sparse case, building on previous results of Janson, Łuczak, and Rucinski allows restriction of the polynomial to the asymptotically contributing portion either when H is connected with non-null 2-core, or when H is a tree. In either case we give a compact computational formula for the asymptotic variance in terms of a rooted tree overlay polynomial. Two cases for which the formula is valid in a range for which both the expected value and variance are finite are when H is a cycle with attached trees and when H is a tree.

Key words: Erdős-Rényi graph, small subgraph, variance, subgraph plot, graph isomorphism
1991 MSC: 05C80, 05C30, 05A15

1 Introduction

The Erdős-Rényi random graph $\mathcal{G}(n,p)$, developed in [1–3], is constructed on the set $[n] := \{1,\ldots,n\}$ of vertices by selecting each potential edge to be present independently with fixed probability p. An historically important

* Author to whom Correspondence should be sent.
Email addresses: rellis@math.iit.edu (Robert B. Ellis), ferry@metsci.com (James P. Ferry).

1 Supported by AFRL contract #FA8750-06-C-0217.

question is the probability threshold function $p(n)$ for the appearance of a copy of a fixed small subgraph H. Deferring most definitions until Section 2, we outline the central role the density of H plays as follows. Define the *maximum density* $m(H) := \{\max(e(F)/v(F)) : F \subseteq H\}$, where F ranges over all subgraphs of H and the *density* $e(F)/v(F)$ is the edge-to-vertex ratio of F. H is *balanced* if $m(H) = e(H)/v(H)$, and otherwise is *unbalanced*. A balanced H is *strictly balanced* if $e(F)/v(F) = m(H)$ only for $F = H$. Erdős and Rényi proved in [3] that $p(n) = n^{-1/m(H)}$ is the threshold for a copy of H appearing when H is balanced, and Bollobás [4] extended this result to unbalanced graphs. Also in [4], and independently by Karoński and Ruciński in [5], the number of copies of H at the threshold was shown to have Poisson distribution when H is strictly balanced. Ruciński and Vince showed in [6] that this in fact is characterizing for strictly balanced H. In [7], Bollobás and Wierman give a subgraph decomposition method for computing the distribution at the threshold for any balanced H, but not a compact formula. We refer the reader to [7–9] for details.

In this paper we take a different approach, calculating the variance of the number of copies of H in $\mathcal{G}(n,p)$. Section 2 presents the necessary definitions and notation, and quotes a formulation of the normalized variance in terms of overlapping copies $\widetilde{H}$ of H. In Section 3, we introduce a graph overlay polynomial $M(H; x, y)$, which is of independent interest, and use it to exactly express the normalized variance in terms of the internal subgraph symmetries of H. This result adapts to arbitrary symmetric random graph processes. A restriction of $M(H; x, y)$ yields the asymptotic variance when $p(n)$ is sufficiently small. Sections 3.1 and 3.2 give compact computational formulas in the cases that H is connected with non-null 2-core, and is a tree, respectively. Along the way we introduce a rooted tree overlay polynomial $B(T, T'; x)$ analogous to $M(H; x, y)$. As a result, we have new compact formulas for the asymptotic variance of the subgraph count at the threshold for a copy of H appearing including two important cases: when H consists of a strictly balanced 2-core (or a sufficiently densely structured core) with trees attached arbitrarily, and when H is a tree. We conclude with several remarks in Section 4.

Studying the variance of the number X_H of copies of H in $\mathcal{G}(n,p)$ is motivated by the following problem of detecting whether H has been inserted into an instance of $\mathcal{G}(n,p)$ by an adversary. Define $\mathcal{G}_H(n,p)$ to be the random graph obtained by pre-inserting a fixed copy of H, and then selecting all remaining edges each independently with probability p. An evidence graph G is presented, but it is unknown whether G was generated from $\mathcal{G}(n,p)$ or $\mathcal{G}_H(n,p)$. The optimal decision statistic from which to choose the most likely generator is the likelihood ratio $\Lambda_H(G)$, which is the ratio of the probability of obtaining G from $\mathcal{G}_H(n,p)$ to the probability of obtaining G from $\mathcal{G}(n,p)$. By Theorem 3 of [10], $\Lambda_H(G) = X_H(G)/E(X(\mathcal{G}(n,p)))$, the ratio of the number of copies of H in G to the expected number of copies of H in $\mathcal{G}(n,p)$. Precise tuning of a decision threshold for the statistic $\Lambda_H(G)$ requires full knowledge of the two distributions for $\Lambda_H(G)$ under the assumption that the generator is $\mathcal{G}(n,p)$ or $\mathcal{G}_H(n,p)$, respectively. A start in the direction of this difficult question is obtaining the variance of $X_H(\mathcal{G}(n,p))$, from which the expected value of $X_H(\mathcal{G}_H(n,p))$ is easily obtained. We refer the reader to [10,11] for details of this detection problem.

2 Definitions and preliminaries

We refer the reader to [9] for a full treatment of the Erdős-Rényi random graph model. Given a nonnegative integer n, a (simple, loopless) graph $G = (V(G), E(G))$ has vertex set $V(G) = [n] := \{1, \ldots, n\}$, and edge set $E(G) \subseteq \binom{[n]}{2} := \{\{i,j\} : i,j \in [n], i \neq j\}$. The complete graph K_n on $[n]$ has edge set $\binom{[n]}{2}$. An edge $e = \{i,j\} \in E(G)$ has *endpoints* $i,j \in V(G)$; we also write $i \sim j$ for vertex adjacency via edge e. When $V(G) \neq \emptyset$, G is *non-null*. The *degree* $d(i)$ of a vertex $i \in V(G)$ is the number of edges for which i is an endpoint. We define $v(G) = n$ and $e(G)$, $0 \leq e(G) \leq \binom{n}{2}$, to be the number of vertices and edges, respectfully, of G. If $V_1, V_2 \subseteq V(G)$ are disjoint, we define $E(V_1, V_2) := \{\{i,j\} \in E(G) : i \in V_1, j \in V_2\}$, with size $e(V_1, V_2) = |E(V_1, V_2)|$, to be the edges of G having one endpoint in each of V_1, V_2. A graph H is a *subgraph* of G, denoted $H \subseteq G$, provided $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. For $U \subseteq V(G)$, $G[U]$ is the subgraph of G *induced* by U, with vertices $V(G[U]) = U$ and edges $E(G[U]) = \{\{i,j\} \in E(G) : i,j \in U\}$. The *graph complement* $H \setminus H_1$ is $H[V(H) \setminus V(H_1)]$. The *intersection* graph $H \cap H_1$ of two subgraphs $H, H_1 \subseteq G$ has vertex set $V(H) \cap V(H_1)$ and edge set $E(H) \cap E(H_1)$. A subgraph $\widetilde{H} \subseteq K_n$ is a *copy* of H provided $\widetilde{H}$ is *isomorphic* to H; that is, there exists a bijective vertex mapping $f : V(H) \to V(\widetilde{H})$ such that $\{i,j\} \in E(H)$ if and only if $\{f(i), f(j)\} \in E(\widetilde{H})$. Define $\mathrm{Iso}(H, \widetilde{H})$ to be the set of all such bijections. If $H = \widetilde{H}$, the isomorphism is an *automorphism*, and we define $\mathrm{Aut}(H) := \mathrm{Iso}(H, H)$.

2.1 The subgraph count and a formula for variance

Given a real number p, $0 \leq p \leq 1$, the Erdős-Rényi random graph $\mathcal{G}(n,p)$ on n vertices is obtained by independently selecting each of $\binom{n}{2}$ possible edges to be present with probability p and absent with probability $1-p$. We may view $\mathcal{G}(n,p)$ as a probability space assigning to each graph G with n vertices probability $p^{e(G)}(1-p)^{\binom{n}{2}-e(G)}$. By the notation $G \sim \mathcal{G}(n,p)$, or "G is distributed as $\mathcal{G}(n,p)$", we mean that G is a random variable selected according to the probability space. $\mathcal{G}(n,p)$ is an example of a *symmetric random graph process*;

that is, one in which isomorphic graphs $H, \widetilde{H}$ have equal probability of being subgraphs of $\mathcal{G}(n,p)$. Let X_H be the number of copies of H in $\mathcal{G}(n,p)$, which is known to have expected value

$$\mathrm{E}(X_H) = \binom{n}{v(H)} \frac{v(H)!}{|\mathrm{Aut}(H)|} p^{e(H)}.$$

The variance of X_H is $\mathrm{Var}(X_H) = E(X_H^2) - E(X_H)^2$, and we define the normalized variance

$$\nu(X_H) := \frac{\mathrm{Var}(X_H)}{\mathrm{E}(X_H)}.$$

We will use the following form of $\nu(X_H)$ due to [12]. For completeness, we include the proof.

Lemma 2.1 *Let H be a fixed graph with vertex set $V(H) \subseteq [n]$. Let $G \sim \mathcal{G}(n,p)$. Then*

$$\begin{aligned} \nu[X_H] &= \sum_{\widetilde{H} \cong H} \left(\mathrm{P}(\widetilde{H} \subseteq G | H \subseteq G) - \mathrm{P}(\widetilde{H} \subseteq G) \right), &(1)\\ &= \sum_{\ell=0}^{e(H)-1} \sum_{\substack{\widetilde{H} \cong H \\ |E(\widetilde{H}) \setminus E(H)| = \ell}} p^{\ell}(1-p) \sum_{i=0}^{e(H)-\ell-1} p^i, &(2) \end{aligned}$$

where additionally (1) *holds for G produced from any symmetric random graph process.*

PROOF. Beginning with the definition of $\nu[X_H] = \mathrm{Var}[X_H]/\mathrm{E}[X_H]$ and assuming $E[X_H] \neq 0$, we have

$$\begin{aligned} \nu[X_H] &= \mathrm{E}[X_H | H \subseteq G] - \mathrm{E}[X_H] \qquad \text{iff}\\ \mathrm{E}[X_H^2] &= \mathrm{E}[X_H | H \subseteq G]\mathrm{E}[X_H] \qquad \text{iff}\\ \sum_{H_1, H_2 \cong H} \mathrm{P}(H_1 \cup H_2 \subseteq G) &= \sum_{H_1 \cong H} \mathrm{P}(H_1 \subseteq G | H \subseteq G) \sum_{H_2 \cong H} \mathrm{P}(H_2 \subseteq G). \qquad (3) \end{aligned}$$

The left-hand side of (3) becomes

$$\sum_{H_1, H_2 \cong H} \mathrm{P}((H_1 \subseteq G) \cap (H_2 \subseteq G)) = \sum_{H_2 \cong H} \sum_{H_1 \cong H} \mathrm{P}(H_1 \subseteq G | H_2 \subseteq G) \mathrm{P}(H_2 \subseteq G). \qquad (4)$$

In a symmetric random graph process such as $\mathcal{G}(n,p)$, isomorphic subgraphs have the same probability. Let $\phi : [n] \to [n]$ be a bijection such that $\phi(H_2) =$

H. Then $\mathrm{P}(H_1 \subseteq G | H_2 \subseteq G) = \mathrm{P}(\phi(H_1) \subseteq G | H \subseteq G)$. Since both H_1 and $\phi(H_1)$ range over all copies of H, we may substitute $\mathrm{P}(H_1 \subseteq G | H \subseteq G)$ into the right-hand side of (4), and factor out the sum over H_2 to obtain the right-hand side of (3). Rewriting $\sum_{H_2 \cong H} \mathrm{P}(H_2 \subseteq G)$ as $\mathrm{E}[X_H]$, dividing by $\mathrm{E}[H_H]$, and subtracting $\mathrm{E}[X_H]$ yields (1). The right-hand side of (1) is equal to $\sum_{\widetilde{H} \cong H} p^{|E(\widetilde{H}) \setminus E(H)|} - p^{e(\widetilde{H})}$, which when refined according to $|E(\widetilde{H}) \setminus E(H)|$ and appropriately factored, becomes (2). □

2.2 *The subgraph plot and leading terms of the variance*

As in Lemma 3.5 of [9], the subgraph count variance can be refined according to $H_1 = \widetilde{H} \cap H$ instead of $|E(\widetilde{H}) \setminus E(H)|$. We quote the result here.

Lemma 2.2 (Janson, Łuczak, Ruciński) *Let H be a fixed graph, and let $G \sim \mathcal{G}(n, p)$. Then the variance of the subgraph count of H in G is*

$$\begin{aligned}\mathrm{Var}(X_H) &= \Theta\left((1-p) \sum_{H_1 \subseteq H, e(H_1)>0} n^{2v(H)-v(H_1)} p^{2e(H)-e(H_1)}\right) \\ &= \Theta\left((1-p) \max_{H_1 \subseteq H, e(H_1)>0} \frac{\mathrm{E}(X_H)^2}{\mathrm{E}(X_{H_1})}\right),\end{aligned} \tag{5}$$

where $\Theta(\cdot)$ indicates asymptotic order of magnitude as $n \to \infty$.

Lemma 2.2 can be obtained from Lemma 2.1 by the mentioned regrouping of terms and by multiplying by $\mathrm{E}(X_H)$ to remove the normalization.

A convenient framework for identifying the dominating terms of (2) (and (5)) is through the subgraph plot of H. The *subgraph plot* of a graph H is the set of points

$$\Sigma(H) := \{(v(H_1), e(H_1)) : H_1 \subseteq H, v(H_1) \geq 2\},$$

which may be visualized in the Cartesian first quadrant $\{(v, e) : v, e \geq 0\}$. We now quote the properties of $\Sigma(H)$ necessary for this paper, and refer the reader to [9] for details. The *roof* of $\Sigma(H)$, denoted by $\hat{\Sigma}(H)$, is the upper boundary of the convex hull of $\Sigma(H)$. We say $H_1 \subseteq H$ lies on the roof if $(v(H_1), e(H_1)) \in \hat{\Sigma}(H)$; H_1 thus has maximum *density* $e(H_1)/v(H_1)$ over all subgraphs of H with $v(H_1)$ vertices. If H_1 lies on the roof, we define $a^+_{H_1}$ to be the (nonnegative) slope of the line segment on the convex hull of $\Sigma(H)$ whose left endpoint is $(v(H_1), e(H_1))$ and whose right endpoint is $(v(H_2), e(H_2)) \in \hat{\Sigma}(H)$ for the next largest possible value of $v(H_2)$. Similarly, $a^-_{H_1}$ is the (nonnegative) slope of the line segment whose right endpoint is $(v(H_1), e(H_1))$ and whose left endpoint is $(v(H_2), e(H_2)) \in \hat{\Sigma}(H)$ for the next smallest possible value of $v(H_2)$. These slopes weakly decrease from the left to the right of $\hat{\Sigma}(H)$. For convenience, define $a_H^+ := 0$ and $a_{K_2}^- := \infty$, where K_2 represents any subgraph of H with 2 vertices and 1 edge. Critically for this paper, $H_1 \subseteq H$ is a *leading term*, or contributes asymptotically to (5), provided $p = p(n)$ simultaneously satisfies $np^{a_{H_1}^+} = \Omega(1)$ and $np^{a_{H_1}^-} = O(1)$. Here, $\Omega(\cdot)$ and $O(\cdot)$ mean asymptotic maximum and minimum order of magnitude, respectively.

2.3 The 2-core and leading terms of the subgraph count variance

The *k-core* of a graph H, defined in [13] and denoted by $C(k; H)$, is the largest subgraph of H with minimum degree at least k. The k-core is unique and equals the graph obtained from H by iterative removal of vertices of degree less than k along with their incident edges. From now on we consider only the 2-core of H, and refer to it as $C(H) := C(2; H)$.

Assume first that $C(H)$ is non-null (and thus $v(C(H)) \geq 3$). In this case define $\hat{\Sigma}_{\mathrm{R}}(H)$ to be the rightmost $v(H) - v(C(H)) + 1$ points on the roof $\hat{\Sigma}(H)$, so that

$$\hat{\Sigma}_{\mathrm{R}}(H) = \{(v(C(H)) + k, e(C(H)) + k) : 0 \leq k \leq v(H) - v(C(H))\}. \quad (6)$$

This is because the edges $E(H) \setminus E(C(H))$ form a forest, and any subgraph $H_1 \subseteq H$ with $v(H_1) \geq v(C(H))$ and maximum density must contain $C(H)$. The result of any breadth-first-search on H with initial graph $C(H)$ gives a sequence of subgraphs achieving (6). Second, if H is a tree, any maximum density subgraph of H is a subtree; in this case define $\hat{\Sigma}_{\mathrm{R}}(H) := \{(2+k, 1+k) : 0 \leq k \leq v(H) - 2\}$. This yields the following characterization of $\hat{\Sigma}_{\mathrm{R}}(H)$ in terms of copies $\widetilde{H}$ of H, by considering $\widetilde{H} \cap H$ to be a partial breadth first search, and filling out $\widetilde{H}$ outside of H.

Lemma 2.3 *If $H \subseteq K_n$ has non-null 2-core $C(H)$ or is a tree with $n \geq 2v(H) - \max(v(C(H)), 2)$, then*

$$\hat{\Sigma}_{\mathrm{R}}(H) = \Big\{(v(\widetilde{H} \cap H), e(\widetilde{H} \cap H)) : \widetilde{H} \subseteq K_n, H \cong H, |E(\widetilde{H}) \setminus E(H)| = \\ |V(\widetilde{H}) \setminus \ V(H)| \in \{0, \ldots, v(H) - \max(v(C(H)), 2)\}\Big\},$$

where $\max(v(C(H)), 2) = 2$ *iff H is a tree.* $\square$

Clearly, for $H_1 \subseteq H$ and $(v(H_1), e(H_1))$ a middle point of $\hat{\Sigma}_{\mathrm{R}}(H)$, $a_{H_1}^+ = a_{H_1}^- = 1$. Moreover, when $C(H)$ is non-null, $a_{C(H)}^+ = a_H^- = 1$; by definition of $\hat{\Sigma}(H)$, $C(H)$ lies on $\hat{\Sigma}(H)$, and

$$a^-_{C(H)} := \min_{\substack{H_1 \subseteq H, \\ 2 \leq v(H_1) < v(C(H))}} \frac{e(C(H)) - e(H_1)}{v(C(H)) - v(H_1)}. \tag{7}$$

From the following result, we have $a^-_{C(H)} > 1$ when H is connected with non-null 2-core, and therefore asymptotic separation in the contribution to $\nu(X_H)$ from $\hat{\Sigma}(H)$ in Lemma 2.1 between when $C(H) \subseteq \widetilde{H} \cap H$ and when $C(H) \nsubseteq \widetilde{H} \cap H$.

Lemma 2.4 *Let H be a graph. If the 2-core $C(H)$ of H is non-null, then $a^-_{C(H)} \geq 1$, with equality iff $C(H)$ has at least two connected components, the least dense of which is a cycle. If H is a tree, then $\hat{\Sigma}_{\mathrm{R}}(H) = \hat{\Sigma}(H)$.*

PROOF. If H is a tree, the result is immediate from Lemma 2.3, so assume $C(H)$ is non-null. Let H_1 achieve (7). We may assume $H_1 = H[V(H_1)]$, since adding only edges to H_1 would reduce $a^-_{C(H)}$. Without loss of generality, replace H_1 with $C(H_1)$, since doing so preserves the order relation (<,>,=) of 1 and $a^-_{C(H)}$. This lets us assume $C(H_1) \subseteq C(H)$. Assume to the contrary that $a^-_{C(H)} < 1$. Then $V(H) \setminus V(H_1) > E(H) \setminus E(H_1)$, forcing a vertex $v \in V(H) \setminus V(H_1)$ with degree 1 in $C(H)$, and contradicting the definition of 2-core. If $a^-_{C(H)} = 1$, then $V(H) \setminus V(H_1) = E(H) \setminus E(H_1)$, and the only way to avoid a degree 1 vertex in $C(H)$ is if all edges in $E(H) \setminus E(H_1)$ are between the vertices of $V(H) \setminus V(H_1)$, giving them average degree 2 in $C(H)$. The only possibility is for $C(H) \setminus C(H_1)$ to be a union of one or more connected components of $C(H)$ which are cycles. Finally, suppose $C(H)$ contains at least two components, one of which is a cycle. Let H_1 be the union of the other components to show that $a^-_{C(H)} \leq 1$. □

Consequently, when $np^{a^-_{C(H)}} = \omega(1)$ and H is connected with non-null 2-core, $\nu(X_H)$ is asymptotically equal to the contribution to (2) when $\widetilde{H} \cap H$ lies on the right side of the roof, $\hat{\Sigma}_{\mathrm{R}}(H)$. Here, $\omega(1)$ means asymptotically strictly greater than any constant function.

3 Exact enumeration of dominating variance terms in the sparse case

In this section we formulate $\nu(X_H)$ in (2) by counting, in terms of subgraph symmetries, the set

$$\mathcal{J}_{k,\ell}(H,n) := \{\widetilde{H} \subseteq K_n : \widetilde{H} \cong H, |V(\widetilde{H}) \setminus V(H)| = k, |E(\widetilde{H}) \setminus E(H)| = \ell\}, \tag{8}$$

for $H \subseteq K_n$ and appropriate k, ℓ. First we define a generating polynomial $M(H; x, y)$ whose coefficients are invariants of H which count symmetries of subgraphs H_1 of H. Theorem 3.2 counts $|\mathcal{J}_{k,\ell}(H,n)|$ in terms of the corresponding coefficient of $M(H; x, y)$, yielding the formula for $\nu(X_H)$ in Cor. 3.3, along with an asymptotic form for the sparse case. We give a computational formula in the sparse case for the main terms of $M(H; x, y)$ in Section 3.1 when H is connected with non-null 2-core, and describe how to adjust it when H is a tree in Section 3.2.

Definition 3.1 *Let H be a graph. Define for $0 \le k \le v(H)$ and $0 \le \ell \le e(H)$ the subgraph collection*

$$\mathcal{H}_{k,\ell}(H) := \{H_1 \subseteq H : |V(H) \setminus V(H_1)| = k, |E(H) \setminus E(H_1)| = \ell\};$$

and for fixed k, ℓ and $H_1, H_2 \in \mathcal{H}_{k,\ell}(H)$ the set of restricted isomorphisms

$$\begin{aligned}\mathrm{Iso}_H(H_1, H_2) := \{\rho \in &\mathrm{Iso}(H_1, H_2) : \forall i, j \in V(H_1), \{i, j\} \in E(H) \setminus E(H_1) \Rightarrow \\ &\{\rho(i), \rho(j)\} \notin E(H) \setminus E(H_2)\};\end{aligned} \tag{9}$$

and the coefficients

$$m_{k,\ell}(H) := \sum_{H_1, H_2 \in \mathcal{H}_{k,\ell}(H)} \frac{|\mathrm{Iso}_H(H_1, H_2)|}{|\mathrm{Aut}(H)|} \tag{10}$$

for the graph overlay polynomial

$$M(H; x, y) := \sum_{k=0}^{v(H)} \sum_{\ell=0}^{e(H)} m_{k,\ell}(H) x^k y^\ell. \tag{11}$$

$\mathcal{H}_{k,\ell}(H)$ is the collection of all $(v(H)-k)$-vertex and $(e(H)-\ell)$-edge subgraphs of H. The set $\mathrm{Iso}_H(H_1, H_2)$ encodes the ways in which an isomorphic copy $\widetilde{H}$ of H can intersect H in H_1, with H_2 being the image of H_1 under an isomorphism from $\widetilde{H}$ to H. The guide for construction of $\widetilde{H}$ is H_2 pulled back under ρ to obtain H_1, which is then extended outside of H to obtain $\widetilde{H}$. Interpreting H_1 as $\widetilde{H} \cap H$, if $\{\rho(i), \rho(j)\} \in E(H) \setminus E(H_2)$, then $\{i, j\}$ must not be in $E(H) \setminus E(H_1)$, since otherwise $\{i, j\}$ is in both $\widetilde{H}$ and H, and thus in H_1. In essence, not every H_1 can be extended to an isomorphic copy of H by using only vertices and edges outside of H. The coefficient $m_{k,\ell}(H)$ counts $|\mathcal{J}_{k,\ell}(H,n)|$ as follows, where $(n - v(H))_k$ equals $\binom{n-v(H)}{k} k!$.

Theorem 3.2 *Let $H \subseteq K_n$ be a graph such that $0 \le k \le v(H)$, $0 \le \ell \le e(H)$, and $n \ge v(H) + k$; and define $\mathcal{J}_{k,\ell}(H,n)$ as in* (8)*. Then*

$$|\mathcal{J}_{k,\ell}(H,n)| = m_{k,\ell}(H)(n - v(H))_k\,.$$

PROOF. Multiplying both sides by $|\mathrm{Aut}(H)|$, it suffices to construct a bijection ϕ from $\{(\widetilde{H},\sigma) \in \mathcal{J}_{k,\ell}(H,n) \times \mathrm{Aut}(H)\}$ to $\{(H_1, H_2, \rho, (x_1,\ldots,x_k)) : H_1, H_2 \in \mathcal{H}_{k,\ell}(H), \rho \in \mathrm{Iso}_H(H_1,H_2), (x_1,\ldots,x_k) \in ([n] \setminus V(H))_k\}$. Here $(X)_k$ is the set of all k-sequences $(x_1,\ldots,x_k)$ with distinct entries from the set X. To define ϕ on $(\widetilde{H},\sigma)$, let $\theta_{\widetilde{H}} \in \mathrm{Iso}(\widetilde{H},H)$ be fixed, depending only on $\widetilde{H}$ and not on σ. For example, $\theta_{\widetilde{H}}$ could be chosen so that when $\{\theta_{\widetilde{H}}(i) : i \in V(\widetilde{H})\}$ is sorted with respect to i, $\theta_{\widetilde{H}}$ gives the lexicographically least list. Let $H_1 = \widetilde{H} \cap H$, let $H_2 = (\sigma \circ \theta_{\widetilde{H}})(H_1)$, and let $\rho = (\sigma \circ \theta_{\widetilde{H}})|_{V(H_1)}$ be the restriction of $\sigma \circ \theta_{\widetilde{H}} \in \mathrm{Iso}(\widetilde{H},H)$ to $V(H_1)$. For all $i,j \in V(H_1)$, if $\{\rho(i),\rho(j)\} \in E(H) \setminus E(H_2)$, then $\{i,j\} \notin E(H) \setminus E(H_1)$; otherwise $\{i,j\}$ lies in both $E(H)$ and in $E(\widetilde{H})$, since $\sigma \circ \theta_{\widetilde{H}}$ preserves edges, and thus $\{i,j\} \in E(H_1)$. Thus the restriction on ρ in (9) is observed. Finally, define $(x_1,\ldots,x_k)$ to be the vertices of $V(\widetilde{H}) \setminus V(H)$ ordered so that $((\sigma \circ \theta_{\widetilde{H}})(x_i) : 1 \le i \le k)$ is in increasing order.

Now assume $(\widetilde{H},\sigma) \neq (\widetilde{H}',\sigma')$ and consider $\phi((\widetilde{H},\sigma)) = (H_1,H_2,\rho,(x_1,\ldots,x_k))$ and $\phi((\widetilde{H}',\sigma')) = (H_1',H_2',\rho',(x_1',\ldots,x_k'))$. In order to show $\phi((\widetilde{H},\sigma)) \neq \phi((\widetilde{H}',\sigma'))$, first suppose $\widetilde{H} = \widetilde{H}'$ but $\sigma \neq \sigma'$; in particular $\theta_{\widetilde{H}} = \theta_{\widetilde{H}'}$ and $H_1 = H_1'$. Additionally assume $H_2 = H_2'$ and $\rho = \rho'$. Since $\sigma \neq \sigma'$, there must exist $x \in V(\widetilde{H}) \setminus V(H)$ such that $(\sigma \circ \theta_{\widetilde{H}})(x) \neq (\sigma' \circ \theta_{\widetilde{H}})(x)$. Then x appears in a different position in $(x_1,\ldots,x_k)$ than in $(x_1',\ldots,x_k')$. Second suppose $\widetilde{H} \neq \widetilde{H}'$. Assume that $H_1 = H_1'$, $H_2 = H_2'$, $\rho = \rho'$, and $\{x_1,\ldots,x_k\} = \{x_1',\ldots,x_k'\}$ (as sets). It remains to show that $(x_1,\ldots,x_k) \neq (x_1',\ldots,x_k')$ (as sequences). As $\widetilde{H} \neq \widetilde{H}'$ and $H_1 = H_1'$, there must exist $y_1,y_2 \in V(\widetilde{H}) = V(\widetilde{H}')$ such that $\{y_1,y_2\} \in E(\widetilde{H}) \setminus E(\widetilde{H}')$. In particular, this forces $((\sigma \circ \theta_{\widetilde{H}})(y_1), (\sigma \circ \theta_{\widetilde{H}})(y_2)) \neq ((\sigma' \circ \theta_{\widetilde{H}'})(y_1), (\sigma' \circ \theta_{\widetilde{H}'})(y_2))$, say due to $(\sigma \circ \theta_{\widetilde{H}})(y_1) \neq (\sigma' \circ \theta_{\widetilde{H}'})(y_1)$, since graph isomorphism preserves edges and non-edges. Since $\rho = \rho'$, y_1 must be in $V(\widetilde{H}) \setminus V(H_1)$ and so its position in $(x_1,\ldots,x_k)$ is distinct from its position in $(x_1',\ldots,x_k')$. Therefore ϕ is one-to-one.

To show ϕ is onto, fix $(H_1,H_2,\rho,(x_1,\ldots,x_k))$. Define $\theta' : V(H_1) \cup \{x_1,\ldots,x_k\} \to V(H)$ as follows. Let $\theta' = \rho$ on $V(H_1)$. Let $\theta'(x_i)$ be the ith lowest element in $V(H) \setminus V(H_2)$. Define $\widetilde{H}$ by letting $\{y_1,y_2\} \in E(\widetilde{H})$ iff $\{\theta'(y_1),\theta'(y_2)\} \in E(H)$, so that $\theta' \in \mathrm{Iso}(\widetilde{H},H)$. In particular $|E(\widetilde{H}) \setminus E(H)| = \ell$, since $\rho \in \mathrm{Iso}_H(H_1,H_2)$. Let $\theta_{\widetilde{H}}$ be the previously determined element of $\mathrm{Iso}(\widetilde{H},H)$, and define σ by $\sigma \circ \theta_{\widetilde{H}} = \theta'$; that is, $\sigma = \theta' \circ \theta_{\widetilde{H}}^{-1} \in \mathrm{Aut}(H)$. Then $\phi(\widetilde{H},\sigma) = (H_1,H_2,\rho,(x_1,\ldots,x_k))$, and ϕ is onto. $\square$

Immediately by Lemmas 2.1, 2.3, and 2.4; by replacing $x^k y^\ell$ in (11) with $(n - v(H))_k (p^\ell - p^{e(H)})$, we have the following.

Corollary 3.3 *Let $H \subseteq K_n$ be a graph, where $n \geq 2v(H)$. Then $\nu(X_H) =$*

$$\sum_{k=0}^{v(H)} \sum_{\ell=0}^{e(H)} \sum_{\widetilde{H} \in \mathcal{J}_{k,\ell}(H,n)} \left(p^\ell - p^{e(H)}\right) = \sum_{k=0}^{v(H)} \sum_{\ell=0}^{e(H)} m_{k,\ell}(H)(n-v(H))_k \left(p^\ell - p^{e(H)}\right) ;$$

furthermore, if H is connected, then the contribution to $\nu(X_H)$ corresponding to $\hat{\Sigma}_{\mathrm{R}}(H)$ is

$$(1+o(1)) \sum_{k=0}^{v(H)-\max(v(C(H)),2)} m_{k,k}(H)(n-v(H))_k \left(p^k - p^{e(H)}\right) , \tag{12}$$

for $p = o(n^{-1/a^-_{C(H)}})$ when $C(H)$ is non-null, and for arbitrary $p = o(1)$ when H is a tree. □

Here, $o(1)$ means asymptotically strictly less than any constant function as $n \to \infty$. The relative order of $m(G)$ and $a^-_{C(H)}$ is arbitrary in general. For a strictly balanced graph, $m(G) < a^-_{C(H)}$, but $m(G) > a^-_{C(H)}$ when $C(H)$ is a large complete graph with a cycle attached. Equality can hold when $C(H)$ is the disjoint union of strictly balanced graphs, such as the union of cycles, by Lemma 2.4.

3.1 When H is connected with non-null 2-core

In view of (12), define $\hat{M}(H;x) := \sum_{k=0}^{v(H)-\max(v(C(H)),1)} m_{k,k}(H)x^k$. When H is a tree, $k = v(H)-1$ is artificially added to the summation for convenience of stating Theorem 3.10; this contributes 0 to $\nu(X_H)$. Our strategy for producing $\hat{M}(H;x)$ is to grow $\widetilde{H}$ by first mapping $C(\widetilde{H})$ onto $C(H)$ using an automorphism, and by then mapping each tree attached to a vertex of $C(\widetilde{H})$ on the tree of its isomorphically corresponding vertex in $C(H)$. We constrain the overall mapping to exclude disconnected $\widetilde{H} \cap H$, but allow all other possible choices for $\widetilde{H} \subseteq K_n$, which comprise $\bigcup_{k \geq 0} \mathcal{J}_{k,k}(H,n)$ (see (8)).

We require some notation for rooted trees and symmetry groups involved in the computation. A *rooted tree* (T,i) is a tree T with a distinguished root vertex $i \in V(T)$, denoted simply by T when the root is clear from context. The null graph $\emptyset$ is neither a tree nor can it be rooted. The *rooted subtrees* of (T,i) are exactly the rooted trees $\{(T(j),j) : j \sim i\}$, where $T(j)$ is the connected component of $T[V(T) \setminus \{i\}]$ containing j. Two rooted trees (T_1,i), (T_2,j) are isomorphic, written $(T_1,i) \cong (T_2,j)$, provided there exists a graph isomorphism $\pi : T_1 \to T_2$ such that $\pi(i) = j$. The group of all such isomorphisms is denoted $\mathrm{Iso(T_1,T_2)}$, and the group of rooted tree automorphisms on T is $\mathrm{Aut}(T) := \mathrm{Iso}(T,T)$.

A *coloring* of $C(H)$ (or more generally any graph) is a mapping $c : V(C(H)) \to [r]$, for some positive integer r. Given $C(H)$ with a coloring c, an automorphism $\pi : C(H) \to C(H)$ is *color-preserving* if for all $i \in V(C(H))$, $c(\pi(i)) = c(i)$. Denote the group of these color-preserving automorphisms by $\mathrm{Aut}_c(C(H))$, which in general is not a normal subgroup of $\mathrm{Aut}(C(H))$. By convention we consider the left cosets $\{\pi \mathrm{Aut}_c(C(H)) : \pi \in \mathrm{Aut}(C(H))\}$ in $\mathrm{Aut}_c(C(H))$, and define $\mathcal{S}_c(C(H))$ to be an arbitrary set of representatives of these left cosets. Removing the edges $E(C(H))$ from H results in a forest of $|V(C(H))|$ trees, each of which we view as being rooted in $V(C(H))$. Thus for each $i \in V(C(H))$, define $T(H, i)$ to be the unique tree in the graph $(V(H), E(H) \setminus E(C(H)))$ containing vertex i, with i as its root. We construct a coloring of $V(C(H))$ based on the types of trees rooted in $V(C(H))$ as follows.

Definition 3.4 Define an equivalence relation on $V(C(H))$ by $i \sim j$ iff $T(H, i) \cong T(H, j)$ for $i, j \in V(C(H))$, and let r be the number of distinct equivalence classes. A *tree-coloring* c of $C(H)$ is any coloring $c : V(C(H)) \to [r]$ such that for all $i, j \in V(C(H))$, $c(i) = c(j)$ iff $T(H, i) \cong T(H, j)$.

The following lemma shows that $\bigcup_{k \geq 0} \mathcal{J}_{k,k}(H, n)$ can be enumerated by partitioning over $\mathcal{S}_c(C(H))$, which is the first step of our computational formula for $\hat{M}(H; x)$.

Lemma 3.5 *Let $H \subseteq K_n$ be connected with non-null 2-core $C(H)$, $0 \leq k \leq v(H) - v(C(H))$, and let $\widetilde{H} \in \mathcal{J}_{k,k}(H, n)$. Let c be a tree-coloring of $C(H)$. Then there exists a unique $\pi \in \mathcal{S}_c(C(H))$ such that $c \circ \pi^{-1}$ is a tree-coloring of $C(\widetilde{H})$ with $T(H, \pi^{-1}(i)) \cong T(\widetilde{H}, i)$ for all $i \in V(C(\widetilde{H}))$.*

PROOF. Let $\theta \in \mathrm{Iso}(H, \widetilde{H})$, and let $\pi_1 = \theta|_{C(H)}$. By design, $\pi_1 \in \mathrm{Aut}(C(H))$, and $T(\widetilde{H}, i) \cong T(H, \theta^{-1}(i))$ for all $i \in V(C(\widetilde{H}))$. Let $\pi_2 \in \mathrm{Aut}_c(C(H))$ such that $\pi := \pi_1 \pi_2 \in \mathcal{S}_c(C(H))$. Then $c \circ \pi^{-1}$ is the desired tree-coloring of $C(\widetilde{H})$. Now let $\pi' \in \mathcal{S}_c(C(H))$ and suppose $c \circ \pi^{-1} = c \circ \pi'^{-1}$. For arbitrary $i \in V(C(\widetilde{H}))$, $\pi^{-1}(i) = j$ and $\pi'^{-1}(i) = j'$ for some $j, j' \in V(C(H))$ with $c(j) = c(j')$. But then $\pi^{-1}\pi'(j') = j$, so that $\pi^{-1}\pi' \in \mathrm{Aut}_c(C(H))$, and π is unique. $\square$

Once $\pi \in \mathcal{S}_c(C(H))$ in Lemma 3.5 is determined, $\widetilde{H}$ is determined by how each $T(\widetilde{H}, i)$ is mapped on $T(H, i)$, and by which vertices of $[n] \setminus V(H)$ fill out the rest of $\widetilde{H}$. These mappings are encoded in the following definition, analogous to $M(H; x, y)$, except that the overlay is required to be connected.

Definition 3.6 *Let $T = (T, i)$ and $T' = (T', j)$ be trees rooted at i and j,* *respectively. Define the* rooted tree overlay polynomial

$$B(T,T';x) := \sum_{(T_1,i)\subseteq T,(T_1',j)\subseteq T'} \frac{|\mathrm{Iso}(T_1,T_1')|}{|\mathrm{Aut}(T')|}\, x^{v(T')-v(T_1')},$$

where the sum is over all rooted subtrees T_1 *and* T_1' *with the same roots as* T *and* T'*, respectively; and the base cases are* $B(\emptyset,T';x) := x^{v(T')}/|\mathrm{Aut}(T')|$ *and* $B(T,\emptyset;x) := 1$.

Theorem 3.7 *Let* H *be a connected graph with non-null 2-core* $C(H)$*, and let* c *be a tree-coloring of* $C(H)$*. Then*

$$\hat{M}(H;x) = \sum_{\pi\in\mathcal{S}_c(C(H))} \prod_{i\in V(C(H))} B(T(H,i),T(H,\pi(i));x)\,, \tag{13}$$

where for $i \in V(C(H))$*, the rooted tree* $T(H,i)$ *is the unique connected component of* $(V(H),E(H)\setminus E(C(H)))$ *containing and rooted at* i*.*

PROOF. As in Definition 3.4, let c be a tree-coloring of $C(H)$. For convenience, define the indexing set $\mathcal{T}_k(H) := \{(T_i \subseteq T(H,i) : i \in V(C(H))) : v(H) - \sum v(T_i) = k\}$ to be the set of all sequences of rooted subtrees with roots in $V(C(H))$ and $v(H)-k$ total vertices; by removing $E(C(H))$ from $E(H)$, these sequences are in bijection with $\mathcal{H}_{k,k}(H)$. Further define $\mathcal{T}(H) := \bigcup_{k=0}^{v(H)-v(C(H))} \mathcal{T}_k(H)$. First multiply $\hat{M}(H;x)$ by $|\mathrm{Aut}(H)|$; a line-by-line justification for each step will follow.

$$\hat{M}(H;x)|\mathrm{Aut}(H)| = \sum_{k=0}^{v(H)-v(C(H))} \sum_{H_1,H_2\in\mathcal{H}_{k,k}(H)} |\mathrm{Iso}(H_1,H_2)|x^k \tag{14}$$

$$= \sum_{k=0}^{v(H)-v(C(H))} \sum_{H_1,H_2\in\mathcal{H}_{k,k}(H)} \sum_{\pi\in\mathrm{Aut}(C(H))} \prod_{i\in V(C(H))} |\mathrm{Iso}(T(H_1,i),T(H_2,\pi(i)))|x^{v(T(H,i))-v(T(H_2,\pi(i)))} \tag{15}$$

$$= \sum_{k=0}^{v(H)-v(C(H))} \sum_{(T_i)\in\mathcal{T}_k(H)} \sum_{(T_i')\in\mathcal{T}_k(H)} \sum_{\pi\in\mathrm{Aut}(C(H))} \prod_{i\in V(C(H))} |\mathrm{Iso}(T_i,T_{\pi(i)}')|x^{v(T(H,i))-v(T_{\pi(i)}')} \tag{16}$$

$$= \sum_{\pi\in\mathrm{Aut}(C(H))} \sum_{(T_i)\in\mathcal{T}(H)} \sum_{(T_i')\in\mathcal{T}(H)} \prod_{i\in V(C(H))} |\mathrm{Iso}(T_i,T_{\pi(i)}')|x^{v(T(H,i))-v(T_{\pi(i)}')} \tag{17}$$

$$= \sum_{\pi\in\mathrm{Aut}(C(H))} \prod_{i\in V(C(H))} \sum_{\substack{T_i\subseteq T(H,i)\\ T_{\pi(i)}'\subseteq T(H,\pi(i))}} |\mathrm{Iso}(T_i,T_{\pi(i)}')|x^{v(T(H,i))-v(T_{\pi(i)}')} \tag{18}$$

$$= |\mathrm{Aut}_c(C(H))| \sum_{\pi \in \mathcal{S}_c(C(H))} \prod_{i \in V(C(H))} |\mathrm{Aut}(T'_{\pi(i)})| \sum_{\substack{T_i \subseteq T(H,i) \\ T'_{\pi(i)} \subseteq T(H,\pi(i))}} \frac{|\mathrm{Iso}(T_i, T'_{\pi(i)})|}{|\mathrm{Aut}(T'_{\pi(i)})|} x^{v(T(H,i)) - v(T'_{\pi(i)})} . \tag{19}$$

Because $|\mathrm{Aut}(H)| = |\mathrm{Aut}_c(C(H))| \prod_i |\mathrm{Aut}(T'_{\pi(i)})|$, and by Definition 3.6, the last line is $|\mathrm{Aut}(H)|$ times the right-hand side of (13). Definition 3.1 immediately gives (14). An element $\rho \in \mathrm{Iso}(\mathrm{H}_1, \mathrm{H}_2)$ is determined by a sequence of choices: restricted to $C(H)$, $\pi = \rho|_{C(H)}$ is an automorphism, and restricted to $T(H_1, i)$, ρ is a rooted tree isomorphism onto $T(H_2, \pi(i))$. Additionally, $k = \sum_{i \in V(C(H))} v(T(H,i)) - v(T(H_2, \pi(i)))$ since $H_2 \in \mathcal{H}_{k,k}(H)$, giving (15). Equation (16) is reached by the previously mentioned definition and properties of $\mathcal{T}_k(H)$. For (17), the sum over k is removed by indexing over $\mathcal{T}(H)$ instead of $\mathcal{T}_k(H)$. When $\sum v(T_i) \neq \sum v(T'_i)$, there is no newly appearing contribution, since $v(T_i) \neq v(T'_{\pi(i)})$ is forced for some i, causing $\mathrm{Iso}(T_i, T'_{\pi(i)}) = \emptyset$. The sum over π is brought outside as it is independent of the choices of T_i, T'_i. A straightforward factorization of the polynomial in terms of the possible contribution of $(T_i, T_{\pi(i)})$ yields (18). To obtain (19), multiply and divide by $\prod_i |\mathrm{Aut}(T'_{\pi(i)})|$. Lemma 3.5 allows us to group $\pi \in \mathrm{Aut}(C(H))$ by left coset representative in $\mathcal{S}_c(C(H))$. When $\pi_1, \pi_2 \in \mathrm{Aut}(C(H))$ are in the same left coset of $\mathrm{Aut}_c(C(H))$, the resulting terms of the summation are identical, since $T(H, \pi_1(i)) \cong T(H, \pi_2(i))$ for all $i \in C(H)$. □

The final step in obtaining a computational formula for $\hat{M}(H; x)$ is to express $B(T, T'; x)$ recursively in terms of rooted subtrees of T and T'. Before presenting the theorem, we require the following definitions to encode mappings of subtrees of T' to those of T, and to group those mappings according to subtree isomorphism.

Definition 3.8 *Let T and T' be rooted trees with roots $r(T)$ and $r(T')$, respectively. Let $T_1, \ldots, T_m$ be the m distinct isomorphism types of the rooted subtrees $T(i)$ of T, and for all $1 \leq a \leq m$, define k_a to be the number of rooted subtrees T isomorphic to T_a. Similarly for T', define $T'_1, \ldots, T'_{m'}$, m', and k'_b, for all $1 \leq b \leq m'$. Define $\mathcal{F}(T, T')$ to be the set of all* overlay functions

$$f : \{i : \{i, r(T)\} \in E(T)\} \to \{j : \{j, r(T')\} \in E(T')\} \cup \{\emptyset\}$$

such that $|f^{-1}(\{j\})| \leq 1$ for $j \neq \emptyset$ (the codomain value $\emptyset$ is formal, and is disjoint from $V(T) \cup V(T')$). Further define the set of all overlay multiplicity vectors $\Gamma(T, T') := \{\gamma(f) : f \in \mathcal{F}(T, T')\}$, *where*

$$\gamma(f) := (\gamma_{ab}(f) : a \in \{0, 1, \ldots, m\}, b \in \{0, 1, \ldots, m'\}, (a, b) \neq (0, 0)),$$

and

$$\gamma_{ab}(f) := \begin{cases} |\{i : f(i) = j, j \neq \emptyset, T(i) \cong T_a, T'(j) \cong T'_b\}|, & a \neq 0 \text{ and } b \neq 0; \\ |\{i : f(i) = \emptyset, T(i) \cong T_a\}|, & a \neq 0 \text{ and } b = 0; \\ |\{j : f^{-1}(\{j\}) = \emptyset, T'(j) \cong T'_b\}|, & a = 0 \text{ and } b \neq 0. \end{cases}$$

Consequently, the elements f of $\mathcal{F}(T, T')$ are partitioned based on $\gamma(f)$.

Theorem 3.9 *Let T and T' be rooted trees, with associated quantities as in Definition 3.8. Then*

$$B(T, T'; x) = \sum_{\gamma \in \Gamma(T,T')} \prod_{a=1}^{m} \left[\binom{k_a}{\gamma_{a1}, \ldots, \gamma_{am'}} \prod_{b=1}^{m'} B(T_a, T_b; x)^{\gamma_{ab}} \right] \prod_{b=1}^{m'} \frac{x^{v(T'_b)\gamma_{0b}}}{\gamma_{0b}! |\mathrm{Aut}(T'_b)|^{\gamma_{0b}}}. \tag{20}$$

PROOF. For convenience of notation, define $\mathcal{U}(T) := \{(T_1(i) \subseteq T(i) : \{i, r(T)\} \in E(T))\}$. Here $T(i)$ is the rooted subtree of T with root $i \sim r(T)$. We abuse notation for this proof only and allow $T_1(i)$ to be either the null graph $\emptyset$, or any rooted tree with root i, which is also a subgraph of $T(i)$. The rooted trees $(T_1, r(T))$ for which $T_1 \subseteq T$ as graphs are in bijection with the sequences of $\mathcal{U}(T)$, since removal of $r(T)$ from T_1 determines the $T_1(i)$'s. We additionally define $|\mathrm{Iso}(\emptyset, \emptyset)| = 1$ and observe that $|\mathrm{Iso}(\emptyset, T)| = |\mathrm{Iso}(T, \emptyset)| = 0$ for any non-null T to allow general decomposition of rooted tree isomorphisms. First multiply $B(T, T'; x)$ by $|\mathrm{Aut}(T')|$ and apply Definition 3.6; a line-by-line justification for each step will follow.

$$|\mathrm{Aut}(T')| B(T, T'; x) = \sum_{\substack{(T_1, r(T)) \subseteq T \\ (T'_1, r(T')) \subseteq T'}} |\mathrm{Iso}(T_1, T'_1)| x^{v(T') - v(T'_1)} \tag{21}$$

$$= \sum_{\substack{(T_1(i)) \in \mathcal{U}(T), \\ (T'_1(j)) \in \mathcal{U}(T')}} \sum_{\substack{f \in \mathcal{F}(T,T'), \\ f(i) = \emptyset \Leftrightarrow T_1(i) = \emptyset, \\ f^{-1}(\{j\}) = \emptyset \Leftrightarrow T'_1(j) = \emptyset}} \left(\prod_{\substack{i \sim r(T), \\ f(i) \neq \emptyset}} |\mathrm{Iso}(T_1(i), T'_1(f(i)))| x^{v(T'(f(i))) - v(T'_1(f(i)))} \right)$$

$$\left(\prod_{\substack{i \sim r(T), \\ f(i) = \emptyset}} |\mathrm{Iso}(T_1(i), \emptyset)| \right) \left(\prod_{\substack{j \sim r(T'), \\ f^{-1}(\{j\}) = \emptyset}} |\mathrm{Iso}(\emptyset, T'_1(j))| x^{v(T'(j))} \right) \tag{22}$$

$$= \sum_{f \in \mathcal{F}(T,T')} \left(\prod_{\substack{i \sim r(T) \\ f(i) \neq \emptyset}} \sum_{\substack{\emptyset \neq T_1(i) \subseteq T(i) \\ \emptyset \neq T'_1(f(i)) \subseteq T'(f(i))}} |\mathrm{Iso}(T_1(i), T'_1(f(i)))| x^{v(T'(f(i))) - v(T'_1(f(i)))} \right)$$

$$\left(\prod_{\substack{j \sim r(T') \\ f^{-1}(\{j\})=\emptyset}} x^{v(T'(j))} \right) \tag{23}$$

$$= \sum_{f \in \mathcal{F}(T,T')} \left(\prod_{\substack{i \sim r(T) \\ f(i) \neq \emptyset}} |\mathrm{Aut}(T'(f(i)))| B\Big(T(i), T'(f(i))\Big) \right)$$

$$\left(\prod_{\substack{j \sim r(T') \\ f^{-1}(\{j\})=\emptyset}} |\mathrm{Aut}(T'(j))| B(\emptyset, T'(j)) \right) \tag{24}$$

$$= \sum_{\gamma \in \Gamma(T,T')} \prod_{a=1}^{m} \left[\binom{k_a}{\gamma_{a1}, \ldots, \gamma_{am'}} \prod_{b=1}^{m'} |\mathrm{Aut}(T'_b)|^{\gamma_{ab}} B(T_a, T'_b; x)^{\gamma_{ab}} \right]$$

$$\prod_{b=1}^{m'} \frac{k'_b!}{\gamma_{0b}!} |\mathrm{Aut}(T'_b)|^{\gamma_{0b}} B(\emptyset, T'_b)^{\gamma_{0b}} \,. \tag{25}$$

This last line is $|\mathrm{Aut}(T')|$ times the right-hand side of (20), by observing that an automorphism on T' is obtained by first permuting isomorphic rooted subtrees in $\prod k'_b!$ ways, and then applying an automorphism to each rooted subtree in $\prod_a \prod_{b \neq 0} |\mathrm{Aut}(T'_b)|^{\gamma_{ab}}$ ways. Equation (22) is obtained by observing that $\rho \in \mathrm{Iso}(T_1, T'_1)$ determines an overlay function $f \in \mathcal{F}(T, T')$ on the rooted subtrees $T_1(i)$ and $T'_1(j)$. Null $T_1(i)$'s and $T'_1(j)$'s must map to null graphs. Factorization yields (23), where the sum over f is brought out front by restricting $T_1(i)$ and $T'_1(j)$ to be null exactly when $f(i) = \emptyset$ and $f^{-1}(\{j\}) = \emptyset$, respectively. Applying Definition 3.6 yields (24). We obtain (25) by counting

$$|\{f : \gamma(f) = \gamma\}| = \prod_{a=1}^{m} \binom{k_a}{\gamma_{a1}, \ldots, \gamma_{am'}} \prod_{b=1}^{m'} \frac{k'_b!}{\gamma_{0b}!}$$

by a routine argument. To this end, fix γ. Construct f by, for each $a = 1, \ldots, m$, apportioning the k_a rooted subtrees $T(i)$ of type (i.e., isomorphism class) T_a into distinct subtypes $T_{a1}, \ldots, T_{am'}$ with multiplicities $\gamma_{a1}, \ldots, \gamma_{am'}$, and with γ_{a0} left over. Now for each $b \neq 0$, Canonically order the rooted subtrees $T(i)$ of subtypes $T_{1b}, \ldots, T_{mb}$; and separately canonically order the k'_b rooted subtrees $T'(j)$ of type T'_b. Permute all k'_b of these rooted subtrees, and un-permute the last γ_{0b} which will not match a $T(i)$. Finally, match each $T(i)$ to a $T'(j)$ by setting $f(i) = j$ starting from the top of each resulting order. $\square$

3.2 When H is a tree

If H is a tree, any isomorphic copy $\widetilde{H} \in \mathcal{J}_{k,k}(H,n)$ intersects H in a subtree. Therefore computing $\hat{M}(H;x)$ amounts to selecting pairs of roots $i, j \in V(H)$, computing rooted tree overlay polynomials $B((H,i),(H,j);x)$, and compensating for over-counting due to multiple choices of rootings leading to a single intersection, and due to the automorphism group on H.

Theorem 3.10 *Let H be a tree, and let $\{[i] \subseteq V(H)\}$ be the set of equivalence classes of $V(H)$ defined by $i_1 \cong i_2$ if there is an automorphism of H sending i_1 to i_2. Then*

$$\hat{M}(H;x) = \left(x^k \to \frac{x^k}{v(H)-k}\right) \sum_{[i],[j]} |[i]| B((H,i),(H,j);x)\,, \tag{26}$$

where (H,i) is H rooted at i, and $\left(x^k \to \frac{x^k}{v(H)-k}\right)$ is the polynomial transformation sending $\sum_j c_j x^j$ to $\sum_j c_j x^j/(v(H)-j)$.

PROOF. First multiply $\hat{M}(H;x)$ by $|\mathrm{Aut}(H)|$; a line-by-line justification for each step will follow.

$$\begin{aligned}
\hat{M}(H;x)|\mathrm{Aut}(H)| &= \sum_{k=0}^{v(H)-1} \sum_{H_1,H_2 \in \mathcal{H}_{k,k}(H)} |\mathrm{Iso}_H(H_1,H_2)| x^k \\
&= \sum_{k=0}^{v(H)-1} \sum_{H_1,H_2 \in \mathcal{H}_{k,k}(H)} \sum_{i \in V(H_1), j \in V(H_2)} \frac{|\mathrm{Iso}((H_1,i),(H_2,j))|}{v(H)-k} x^k && (27)\\
&= \sum_{i,j \in V(H)} \sum_{(H_1,i) \subseteq H, (H_2,j) \subseteq H} \frac{|\mathrm{Iso}((H_1,i),(H_2,j))|}{v(H)-k} x^{v(H)-v(H_2)} && (28)\\
&= \sum_{i,j \in V(H)} |\mathrm{Aut}((H,j))| \left(x^k \to \frac{x^k}{v(H)-k}\right) B((H,i),(H,j);x) && (29)
\end{aligned}$$

Equation (27) is reached by identifying each graph isomorphism $\rho \in \mathrm{Iso}_H(H_1,H_2)$ with $v(H_1) = v(H) - k$ rooted tree isomorphisms; letting $i \in H_1$ and setting $j = \rho(i) \in H_2$ yields an element of $\mathrm{Iso}((H_1,i),(H_2,\rho(i)))$. To obtain (28), the sum over k is removed by letting H_1 and H_2 range over all subtrees of H, and the sum over i and j brought out front by rooting H_1 at i and H_2 at j. Applying Definition 3.6 yields (29). To obtain the right-hand side of (26), note that $|\mathrm{Aut}(H)| = |[j]||\mathrm{Aut}((H,j))|$, since an automorphism of H permutes j within its equivalence class and then applies a rooted tree automorphism of $|\mathrm{Aut}((H,j))|$. □

4 Concluding remarks

In the authors' opinion, the most interesting applications of the formulas for $\hat{M}(H;x)$ are direct computation of the asymptotic variance of X_H at the threshold $p = cn^{-1/m(H)}$ (i) when H is a cycle with trees attached arbitrarily by their roots (balanced with strictly balanced 2-core), and (ii) when H is a tree. In both cases the distribution of X_H has finite expectation and finite variance. Additional cases can be analyzed by restriction starting from Corollary 3.3. The variance formula of (12) is asymptotically correct but tends to infinity when the 2-core of H is denser than a cycle and p is such that $0 < \mathrm{P}(X_H > 0) < 1$. The usefulness of the symmetry-infused graph and rooted tree overlay polynomials $M(H;x,y)$ and $B(T,T';x)$, which are valid independent of the random graph process, bears further investigation, as does the derivation of better estimates for the variance within individual families of subgraphs.

Acknowledgments

We thank Darren Lo for several helpful comments, and the anonymous referees for their careful review of the manuscript.

References

[1] P. Erdős, Some remarks on the theory of graphs, Bull. Amer. Math. Soc. 53 (1947) 292–294.

[2] P. Erdős, A. Rényi, On random graphs. I, Publ. Math. Debrecen 6 (1959) 290–297.

[3] P. Erdős, A. Rényi, On the evolution of random graphs, Magyar Tud. Akad. Mat. Kutató Int. Közl. 5 (1960) 17–61.

[4] B. Bollobás, Threshold functions for small subgraphs, Math. Proc. Cambridge Philos. Soc. 90 (2) (1981) 197–206.

[5] M. Karoński, A. Ruciński, On the number of strictly balanced subgraphs of a random graph, in: Graph theory (Łagów, 1981), Vol. 1018 of Lecture Notes in Math., Springer, Berlin, 1983, pp. 79–83.

[6] A. Ruciński, A. Vince, Balanced graphs and the problem of subgraphs of random graphs, Congr. Numer. 49 (1985) 181–190.

[7] B. Bollobás, J. C. Wierman, Subgraph counts and containment probabilities of balanced and unbalanced subgraphs in a large random graph, in: Graph theory

and its applications: East and West (Jinan, 1986), Vol. 576 of Ann. New York Acad. Sci., New York, 1989, pp. 63–70.

[8] B. Bollobás, Random graphs, 2nd Edition, Cambridge University Press, Cambridge, 2001.

[9] S. Janson, T. Łuczak, A. Rucinski, Random graphs, Wiley-Interscience, New York, 2000.

[10] T. L. Mifflin, C. Boner, G. A. Godfrey, J. Skokan, A random graph model for terrorist transactions, in: IEEE Aerospace Conference, 2004, Vol. 5, 2004, pp. 3258–3264.

[11] J. P. Ferry, D. Lo, S. T. Ahearn, A. M. Phillips, Network detection theory, in: N. Memon, J. D. Farley, D. L. Hicks, T. Rosenorn (Eds.), Mathematical Methods in Counterterrorism, Springer-Verlag Vienna, 2009, pp. 161–181.

[12] S. Ahearn, A. Phillips, private communication.

[13] S. B. Seidman, Network structure and minimum degree, Social Networks 5 (3) (1983) 269–287.