

JOINT DEGREE DISTRIBUTIONS OF PREFERENTIAL ATTACHMENT RANDOM GRAPHS

EROL PEKÖZ,* *Boston University*

ADRIAN RÖLLIN,** *National University of Singapore*

NATHAN ROSS,*** *University of Melbourne*

Abstract

We study the joint degree counts in linear preferential attachment random graphs and find a simple representation for the limit distribution in infinite sequence space. We show weak convergence with respect to the p -norm topology for appropriate p and also provide optimal rates of convergence of the finite-dimensional distributions. The results hold for models with any general initial seed graph and any fixed number of initial outgoing edges per vertex; we generate nontree graphs using both a lumping and a sequential rule. Convergence of the order statistics and optimal rates of convergence to the maximum of the degrees is also established.

Keywords: Preferential attachment random graph; multicolor Pólya urn; distributional approximation

2010 Mathematics Subject Classification: Primary 60F05; 60C05
Secondary 05C80; 60K99

1. Introduction

Preferential attachment random graph models have become extremely popular in the fifteen years since they were studied by Barabási and Albert [4]. In the basic models, nodes are sequentially added to the network over time and connected randomly to existing nodes such that connections to higher degree nodes are more likely. The literature around these models has become too vast to survey, but Newman [17], Newman *et al.* [18], and van der Hofstad [28] provide good overviews.

The most popular models are those similar to Barabási and Albert [4], in which nodes are added sequentially and attach to exactly one randomly chosen existing node, and the chance a new node connects to an existing node is proportional to its degree; see [8], [14], and [26] for results on more general attachment rules. The model is typically generalized to allow for vertices to have $\ell \geq 1$ initial edges by starting with the previous model and then, for $k = 0, 1, 2, \dots$, lumping vertices $k\ell + 1, k\ell + 2, \dots, (k + 1)\ell$ into a single vertex (possibly causing loops). The most studied feature of these objects is the distribution of the degrees of the nodes; that is, the proportion of nodes that have degree k as the graph grows large. The basic content of Barabási and Albert [4] and the rigorous formulation of Bollobás *et al.* [6] is that, as $k \rightarrow \infty$,

Received 15 March 2016; revision received 21 September 2016.

* Postal address: Questrom School of Business, Boston University, 595 Commonwealth Avenue, Room 607, Boston, MA 02215, USA. Email address: pekoz@bu.edu

** Postal address: Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore 117546, Singapore. Email address: adrian.roellin@nus.edu.sg

*** Postal address: School of Mathematics and Statistics, University of Melbourne, Richard Berry Building, VIC 3010, Australia. Email address: nathan.ross@unimelb.edu.au

this distribution roughly decays proportional to $k^{-\gamma}$ for some $\gamma > 0$; this is the so-called power law behavior.

In this paper we study the joint degree distribution for a linear preferential attachment model where each entering node initially attaches to exactly $\ell \geq 1$ nodes. In the $\ell \geq 2$ case, we study two mechanisms for attaching edges, a sequential update rule and the lumping rule mentioned above (and note our results are strikingly different for the two rules). We show weak convergence with respect to p -norm topology of the scaled degrees for the process started from any initial ‘seed’ graph to a limiting distribution that has a simple representation and provide an optimal rate of convergence for the finite-dimensional distributions.

To state our results in greater detail, we first precisely define the random rules governing the evolution of the models we study. We distinguish between two cases: one that allows for loops and the other that does not; in fact, the two can be related (see Lemma 1 below), but we present the results separately for the sake of clarity.

Our results are stated in terms of weights of vertices, but our weights can be thought of as in-degree plus one since in the models, vertices are born with weight 1, and each time a vertex receives a new edge from another vertex, its weight increases by 1. Fix $\ell \geq 1$ and let $d_k(n)$ denote the weight of vertex k in the graph $G(n)$. Assume that the seed graph $G(0)$ has s vertices with labels $1, \dots, s$ and with initial weights d_1, \dots, d_s (note that $d_i = d_i(0)$). We construct the graph $G(n)$ with n vertices from $G(n - 1)$ having $n - 1$ edges in two possible ways.

1.0.1. *Model N_ℓ .* Given the graph $G(n - 1)$ having $s + n - 1$ vertices, $G(n)$ is formed by adding a vertex labelled $s + n$ and sequentially attaching ℓ edges between it and the vertices of $G(n - 1)$ according to the following rules. The first edge attaches to vertex k with probability

$$\frac{d_k(n - 1)}{\sum_{i=1}^{s+n-1} d_i(n - 1)}, \quad 1 \leq k \leq n - 1;$$

denote by K_1 the vertex which received that first edge. The weight of K_1 is updated immediately, so that the second edge attaches to vertex k with probability

$$\frac{d_k(n - 1) + \mathbf{1}(k = K_1)}{1 + \sum_{i=1}^{s+n-1} d_i(n - 1)}, \quad 1 \leq k \leq n - 1,$$

where $\mathbf{1}(\cdot)$ is the indicator function. The procedure continues this way, edges attach with probability proportional to weights at that moment, and additional received edges add one to the weight of a vertex, until vertex n has ℓ outgoing edges. Finally, we set $d_{s+n}(n) = 1$, and let $G(n)$ be the resulting graph. Note that multiple edges between vertices are possible.

1.0.2. *Model L_ℓ .* Given the graph $G(n - 1)$ having $s + n - 1$ vertices, $G(n)$ is formed by adding a vertex labelled $s + n$ having weight $d_{s+n}(n - 1) = 1$ and attaching ℓ edges between it and the vertices labelled $\{1, \dots, s + n\}$ (so loops are possible). The first edge attaches to vertex k with probability

$$\frac{d_k(n - 1)}{\sum_{i=1}^{s+n} d_i(n - 1)}, \quad 1 \leq k \leq n;$$

denote by K_1 the vertex which received that first edge. As in model N_ℓ , the weight of K_1 is updated immediately, and the procedure continues in this way until ℓ edges are added and we call $G(n)$ the resulting graph.

Note that if $\ell = 1$ then the weights can be interpreted as the total degree since each additional vertex has ‘out-degree’ 1 and, in this case, the N_1 model is just the usual Barabási–Albert preferential attachment tree.

In the next section we state our finite-dimensional results; process level statements are in Section 1.2.

1.1. Finite-dimensional degree distributions

Key to both understanding and obtaining our results is to focus on the joint cumulative degree counts rather than the joint degree counts. Write $X \sim GG(a, b)$ with $a > 0$ and $b > 0$ for a random variable X having the generalized gamma distribution with density proportional to $x^{a-1}e^{-x^b}$ on $x > 0$, and $Y \sim \text{beta}(a, b)$ for $a > 0$ and $b > 0$ if Y has density proportional to $x^{a-1}(1-x)^{b-1}$ on $0 < x < 1$.

Theorem 1. Fix a seed graph $G(0)$ having s vertices and weight sequence d_1, \dots, d_s , and let $m_k = \sum_{i=1}^k d_i$ for $1 \leq k \leq s$. Assume that either

- (i) $G(n)$ follows model N_ℓ , in which case let $a_k = m_s + (\ell + 1)(k - s) + \ell$ for $k \geq s$; or
- (ii) $G(n)$ follows model L_ℓ , in which case let $a_k = m_s + (\ell + 1)(k - s)$ for $k \geq s$.

Fix $r \geq s$ and let B_1, \dots, B_{r-1} and Z_r be independent random variables with distributions

$$B_k \sim \begin{cases} \text{beta}(m_k, d_{k+1}) & \text{if } 1 \leq k < s, \\ \text{beta}(a_k, 1) & \text{if } s \leq k \leq r, \end{cases}$$

and $Z_r \sim GG(a_r, \ell + 1)$. Define the products

$$Z_k = B_k \dots B_{r-1} Z_r, \quad 1 \leq k < r,$$

and let $Z = (Z_1, \dots, Z_r)$ and $Y = (Z_1, Z_2 - Z_1, \dots, Z_r - Z_{r-1})$. Denote the scaled weight sequence of the first r vertices of $G(n)$ by

$$D(n) = (D_1(n), \dots, D_r(n)) = \frac{1}{(\ell + 1)n^{\ell/(\ell+1)}}(d_1(n), \dots, d_r(n)).$$

Then there is a positive constant $C = C(r, \ell, m_s)$ such that

$$\sup_K |\mathbb{P}[D(n) \in K] - \mathbb{P}[Y \in K]| \leq \frac{C}{n^{\ell/(\ell+1)}} \quad \text{for all } n \geq 1,$$

where the supremum ranges over all convex subsets $K \subset \mathbb{R}^r$.

Remark 1. The error rate $n^{-\ell/(\ell+1)}$ is the best possible since the rate of convergence of a scaled integer-valued random variable to a limiting distribution with nice density is bounded from below by the scaling (nice means uniformly bounded away from 0 on some interval; see [19, Lemma 4.1]). Also, a bound on the constant $C(r, \ell, m_s)$ could, in principle, be made explicit with our methods, but with much added technicality. Such a constant would increase in each of its arguments. We emphasize that obtaining optimal rates in multidimensional distributional limit theorems for the metric we are using is, in general, neither easy nor common.

Since the sets $\{(x_1, \dots, x_r) : \max\{x_1, \dots, x_r\} \leq t\}$ are convex in \mathbb{R}^r , we immediately obtain the following corollary to the theorem.

Corollary 1. Let D and Y be as in Theorem 1 under either (i) or (ii). Then, there is a positive constant $C = C(r, \ell, m_s)$, such that

$$\sup_{t \geq 0} \left| \mathbb{P}\left[\max_{1 \leq k \leq r} D_k(n) \leq t\right] - \mathbb{P}\left[\max_{1 \leq k \leq r} Y_k \leq t\right] \right| \leq \frac{C}{n^{\ell/(\ell+1)}} \quad \text{for all } n \geq 1.$$

The representation of the limits appearing in Theorem 1 is a sort of ‘backward’ construction started from Z_r . Our next result is an alternative ‘forward’ construction that is useful for describing the infinite limit $\mathcal{L}(Z_1, Z_2, \dots)$; compare to the forthcoming process level discussion. Denote by $\text{gamma}(r, \lambda)$ a gamma distribution with shape parameter $r > 0$ and rate parameter $\lambda > 0$ having density proportional to $x^{r-1}e^{-\lambda x}$, $x > 0$.

Proposition 1. *Let B_1, \dots, B_{s-1} and a_s be as in Theorem 1, and let X_1, X_2, \dots be independent with $X_1 \sim \text{gamma}(a_s/(\ell + 1), 1)$ and $X_k \sim \text{gamma}(1, 1)$, $k \geq 2$. Then the random vector Z in Theorem 1 has the representation $Z \stackrel{D}{=} \tilde{Z}$, where*

$$\tilde{Z}_k = \begin{cases} B_k \dots B_{s-1} X_1^{1/(\ell+1)} & \text{if } 1 \leq k < s, \\ (X_1 + \dots + X_{k-s+1})^{1/(\ell+1)} & \text{if } k \geq s. \end{cases}$$

Proof. We need to check that the representation above has the same distribution as that of Theorem 1. The two representations are identical for $k = 1, \dots, s - 1$. For $k \geq s$, that the joint distributions continue to agree is an easy consequence of induction and the beta-gamma algebra which implies that, for $k > s$,

$$(X_1 + \dots + X_{k-s})^{1/(\ell+1)} = \left(\frac{X_1 + \dots + X_{k-s}}{X_1 + \dots + X_{k-s+1}} \right)^{1/(\ell+1)} (X_1 + \dots + X_{k-s+1})^{1/(\ell+1)} \\ \stackrel{D}{=} V^{1/(\ell+1)} Z_k,$$

where $V \sim \text{beta}(a_s/(\ell + 1) + k - s - 1, 1)$ is independent of Z_k . A simple calculation shows that $V^{1/(\ell+1)} \sim \text{beta}(a_s + (\ell + 1)(k - s - 1), 1)$, which is the same distribution as B_{k-1} . Continuing in this way yields the proposition. □

Remark 2. Proposition 1 leads to rather clean representations of the limits appearing in Theorem 1 for particular choices of a seed graph. If the seed graph $G(0)$ has one vertex with $d_1 = \ell + 1$ and $G(n)$ is formed according to model L_ℓ (alternatively, the seed graph has two vertices with $d_1 = \ell + 1$ and $d_2 = 1$, following model N_ℓ), then Proposition 1 implies that $0 < Z_1 < Z_2 < \dots$ are the points of an inhomogeneous Poisson point process on the positive line with intensity $(\ell + 1)t^\ell dt$.

1.2. Process level convergence and order statistics

Using Kolmogorov’s extension theorem, for any choice of seed graph and generative mechanism, the distribution of the vector Y in Theorem 1 can be uniquely extended to obtain an infinite vector $\tilde{Y} = (Y_1, Y_2, \dots)$ which we can take to be of the form $Y_i = Z_i - Z_{i-1}$ for $0 < Z_1 < Z_2 < \dots$ given explicitly by the forward construction of Proposition 1. It is immediate from Theorem 1 that $\tilde{D}(n) = (D_1(n), \dots, D_n(n), 0, \dots)$, viewed as an infinite sequence by appending zeros, converges weakly to \tilde{Y} with respect to the product topology. However, weak convergence with respect to this topology does not yield much at the process level (for example, restricting to bounded sequences, it does not imply convergence of the sequence of maximums to the maximum of the limit), and so we show weak convergence with respect to a topology that is strong enough to imply weak convergence of order statistics.

For $1 \leq p < \infty$, let l_p be the usual sequence-space endowed with the norm $\|\cdot\|_p$. Moreover, denote by $l_p^+ \subset l_p$ the subspace of sequences having nonnegative entries along with the topology inherited from l_p . Clearly, $\tilde{D}(n) \in l_p^+$ for all $n \geq 1$. For $x \in l_p^+$, denote by x^\downarrow the sequence with the entries of x appearing in decreasing order; clearly $x^\downarrow \in l_p^+$.

We have the following result which greatly generalizes the convergence of the maximum of the first r coordinates of $D(n)$ in Corollary 1 to the order statistics of the entire sequence $\tilde{D}(n)$.

Theorem 2. *Let \tilde{Y} and $\tilde{D}(n)$ be the infinite extensions of the random vectors Y and $D(n)$ of Theorem 1 as just described. Then $\tilde{Y} \in l_p^+$ almost surely for any $p > (\ell + 1)/\ell$. Moreover, for $\ell = 1$ and any $p \geq 4$, or $\ell \geq 2$ and any $p \geq \ell + 1$, the sequences $\mathcal{L}(\tilde{D}(n))$ and $\mathcal{L}(\tilde{D}(n)^\downarrow)$ converge weakly to $\mathcal{L}(\tilde{Y})$ and $\mathcal{L}(\tilde{Y}^\downarrow)$ with respect to the l_p -topology.*

1.3. Idea of the proofs of Theorem 1 and Theorem 2

The representation of the limit vector in Theorem 1 is integral to our approach. It says that in the limit for $k > s$, conditional on coordinate Z_{k+1} , the previous coordinate Z_k is an independent beta variable multiplied by Z_{k+1} . On the other hand, for $k > s$, conditional on the sum of the weights of the first $k + 1$ vertices, say $S_{k+1}(n)$, it is not too difficult to see that the sum of the weights of the first k vertices, $S_k(n)$, will be distributed as a classical Pólya urn, run for a number of steps of the order $S_{k+1}(n)$ (see Lemma 2 below). Furthermore, Pólya urns limit to beta variables, so $\mathcal{L}(S_k(n)) \approx \mathcal{L}(B_k S_{k+1}(n))$, where B_k is an appropriate beta variable independent of $S_{k+1}(n)$ (see Lemma 3 below, an extension of [21, Lemma 4.4], which gives a new bound on the Wasserstein distance between Pólya urns and beta distributions). The limits satisfy this approximate identity exactly; that is, $Z_k = B_k Z_{k+1}$. Once we take care of the error made in swapping out Pólya urns for betas (done via a telescoping sum argument), all that is left is to show that Z_r is close to $S_r(n)$, which follows by the results of Peköz *et al.* [21]. For Theorem 2, we establish weak convergence of the distribution of the sequence $\tilde{D}(n)$ to the distribution of \tilde{Y} with respect to l_p topology by verifying a tightness criterion for probability measures on l_p , and then applying Theorem 1 to show the convergence of finite-dimensional distributions. The convergence of the distribution of $\tilde{D}^\downarrow(n)$ follows from the continuous mapping theorem applied to the ordering function.

1.4. Related work

To the best of the authors’ knowledge all of the results above for $\ell \geq 2$ are new. The sequential-edge model was studied by Berger *et al.* [5], where they gave a beautiful description of the limiting structure of the graph started from a random vertex and performing a depth first search. That work is complementary to ours: as the number of vertices $n \rightarrow \infty$, the depth first search started from a uniformly chosen vertex only sees vertices of order n , and these do not appear in our limits.

For $\ell = 1$, the limiting marginal distribution of the degree of the i th vertex are described in different ways in [13] (after relating these variables to an appropriate triangular urn model) and [19, Theorem 1.1, Proposition 2.3]. In the latter paper models 1 and 2 are our models N_1 with $s = 2$ and $d_1 = d_2 = 1$, and L_1 with $s = 1$ and $d_1 = 2$, respectively, and their scaling is a constant times \sqrt{n} . For model N_1 , they identify the distributional limit of $D_i(n)$ for $i \geq 2$ (note that vertices 1 and 2 have the same distribution in this model) as $(\Gamma_1 B_{1/2, i-3/2})^{1/2}$, where $\Gamma_1 \sim \text{gamma}(1, 1)$ is independent of $B_{1/2, i-3/2} \sim \text{beta}(\frac{1}{2}, i - \frac{3}{2})$. In model L_1 , the analogous limit can be written as $(\Gamma_1 B_{1/2, i-1})^{1/2}$ (interpreting $B_{1/2, 0} = 1$). Using our description of these limits given by Proposition 1, we find that, for $X_{1/2} \sim \text{gamma}(\frac{1}{2}, 1)$ independent of X_1, X_2, \dots , independent and identically distributed (i.i.d.) random variables with distribution $\text{gamma}(1, 1)$,

$$\begin{aligned} \sqrt{X_{1/2} + X_1 + \dots + X_{i-1}} - \sqrt{X_{1/2} + X_1 + \dots + X_{i-2}} &\stackrel{D}{=} \sqrt{\Gamma_1 B_{1/2, i-3/2}}, \\ \sqrt{X_1 + X_2 + \dots + X_i} - \sqrt{X_1 + X_2 + \dots + X_{i-1}} &\stackrel{D}{=} \sqrt{\Gamma_1 B_{1/2, i-1}}. \end{aligned}$$

These intriguing distributional identities are not easily interpretable, but they can be directly verified by comparing Mellin transforms. It would be of interest to obtain a representation of the joint distributions similar in appearance to the right-hand side of these identities.

Still assuming that $\ell = 1$, there are some existing results about properties and characterizations of limits of the joint degrees of fixed vertices and maximums of such, but only started from certain seed graphs. Especially close to our work in this special case is Móri [16], who used martingale arguments to show that $(2\tilde{D}(n))_{n \geq 1}$, in model N_1 with $s = 2$ and $d_1 = d_2 = 1$, has an almost sure limit denoted by $(\zeta_1, \zeta_2, \dots)$, which must have the same distribution as $2\tilde{Y}$, though the description of the limiting ζ_i previously given is not so explicit: moment formula are given (and can be checked to agree with those of $2\tilde{Y}$) and some other properties are derived. For example, Móri [16, Lemma 3.4 with $\beta = 0$] showed that the variables

$$\tau_j := \frac{\zeta_1 + \dots + \zeta_{j-1}}{\zeta_1 + \dots + \zeta_j}$$

are $\text{beta}(2j - 1, 1)$ and that $\tau_1, \dots, \tau_r, \zeta_1 + \dots + \zeta_r$ are independent. This coincides with our description of the limits in Theorem 1. See also [12, Section 2] for a discussion of these representations in this and related models. Again using martingales, Móri [16, Theorem 3.1] showed that $(\max_{i \geq 1} D_i(n))_{n \geq 1}$ converges almost surely and in L_p for $p \geq 1$ to $\max_{i \geq 1} \zeta_i$. Thus, we can identify the distribution of this limit as that of $2 \max_{i \geq 1} Y_i$.

Our work generalizes and extends these existing results for the $\ell = 1$ case in several directions. The rates of convergence in Theorem 1 and Corollary 1 are new, as are the descriptions of the limiting joint degrees, even for simple seed graphs where existing results are available. At the process level, our L_p convergence complements the almost sure convergence of the sequence and maximum given by Móri [16] for the basic seed graph, as well as covering much more, allowing different seed graphs and giving convergence of the order statistics.

1.4.1. *A different multiedge preferential attachment graph.* An alternative way to define a preferential attachment model where each new node attaches to $\ell > 1$ nodes was given by Bollobás *et al.* [6]. The model begins by generating a random graph according to model L_1 or N_1 with $n\ell$ nodes, denoted $G(n\ell)$, and then, for each of $i = 1, \dots, n$, collapsing nodes $(i - 1)\ell + 1, \dots, i\ell$ into one node keeping all of the edges (so there may be loops in both models). But with this definition the degree of the i th node is just the sum of the degrees of nodes $(i - 1)\ell + 1, \dots, i\ell$ in $G(n\ell)$ and so the finite-dimensional limits can be read from Theorems 1. Moreover, since a linear transformation of a convex set is convex, the analogous error rates of the theorem in this more general setting also hold. An important remark is that this multiedge preferential attachment graph is fundamentally different than that introduced above: in this model the degree of a fixed vertex grows like \sqrt{n} for any ℓ where, as in models N_ℓ and L_ℓ , the degree grows like $n^{\ell/(\ell+1)}$. Also note that using the representation of Proposition 1, summing the limiting distributions of the degrees of adjacent vertices has a particularly simple form due to telescoping. For example, for model L_1 started from a single loop (so $s = 1$ and $d_1 = 2$), the joint distributional limits of the scaled degrees of nodes $i = 1, \dots, r$ in $G(n)^{(\ell)}$ are given by

$$\sqrt{X_1 + \dots + X_{\ell i}} - \sqrt{X_1 + \dots + X_{\ell(i-1)}},$$

where X_1, X_2, \dots are i.i.d. and have distribution $\text{gamma}(1, 1)$.

Similarly, at the process level, the lumping operation maps L_p^+ to L_p^+ since by Hölder’s inequality, for numbers x_1, \dots, x_ℓ ,

$$|x_1 + \dots + x_\ell|^p \leq \ell^{p-1}(|x_1|^p + \dots + |x_\ell|^p).$$

The lumping is also Lipschitz continuous since, for x, y in l_p^+ ,

$$\sum_{i \geq 1} \left| \sum_{j=1}^{\ell} x_{(i-1)\ell+j} - \sum_{j=1}^{\ell} y_{(i-1)\ell+j} \right|^p \leq \sum_{i \geq 1} \sum_{j=1}^{\ell} |x_{(i-1)\ell+j} - y_{(i-1)\ell+j}|^p.$$

So by the continuous mapping theorem, we can easily read the limits of the degrees for the lumped graph from those of the original graph given in Theorem 2. We do not provide a formal statement of our results for this case because it is a straightforward derivative of the $\ell = 1$ case.

1.4.2. *Connection to the Brownian continuum random tree.* Consider model L_1 started from a loop (so $s = 1$ and $d_1 = 2$). If we write $S_i(n) = \sum_{j=1}^i D_j(n)$ then Proposition 1 and the results of Móri [16], discussed above, imply that, for $r \geq 1$, the scaled sums of degree counts

$$(S_1(n), \dots, S_r(n)) \xrightarrow{\text{a.s.}} (\sqrt{X_1}, \sqrt{X_1 + X_2}, \dots, \sqrt{X_1 + \dots + X_r}),$$

where the X_i are i.i.d. and have distribution $\text{gamma}(1, 1)$. These are the points of an inhomogeneous Poisson point process with intensity $2t \, dt$, which also arises in Aldous' continuum random tree (CRT) construction (see [1] and [2]), and is described by Pitman [24, around Theorem 7.9]. The explicit connection is that if we consider $G(n)$ plus the not yet attached half-edge of vertex $n + 1$, then there are $2n + 1$ 'degrees' which can be bijectively mapped to a binary tree with $n + 1$ leaves. The bijection is defined through Rémy's algorithm for generating uniformly chosen binary plane trees (see the discussion in [21, Remark 2.6]). The algorithm begins with a binary tree with two leaves and a root, corresponding to the two starting degrees of the loop and the half-edge of the second vertex in model L_1 . In Rémy's algorithm, leaves are added to the tree by selecting a (possibly internal) vertex uniformly at random and inserting a cherry at this vertex (that is, insert a graph with three vertices and two edges with the 'elbow' oriented towards the root of the binary tree), so two vertices (one of which is a leaf) are added at each step and these correspond to the two degrees of each edge added in the preferential attachment model. The number of vertices in the spanning tree of the first k leaves added in Rémy's algorithm is exactly the sum of the degrees of the first k vertices in model L_1 started from a loop. If the leaves are labelled in the order they appear (the initial two leaves labelled 1 and 2), then this leaf labelling is uniform which implies that if we first choose a uniform random binary plane tree with $n \geq 2$ leaves and then $k \leq n$ leaves uniformly at random and fix a labelling $1, \dots, k$, then for $T_j(n)$ defined to be the number of vertices in the spanning tree containing the root and the leaves labelled $1, \dots, j$, we have

$$(S_1(n), \dots, S_k(n)) \stackrel{D}{=} \frac{1}{2n^{1/2}} (T_1(n), \dots, T_k(n)).$$

Theorem 1 provides a rate of convergence of this random vector to its limit.

We can now clearly see the connection to the CRT since according to Aldous [2], uniform random binary plane trees converge to Brownian CRT, and the number of vertices in the spanning tree of k randomly chosen leaves in a uniform binary tree of n leaves converges to the length of the tree induced by Brownian excursion sampled at k uniform times as per Pitman [23], [24, Theorem 7.9], and it is known that these trees are formed by combining branches with lengths given by a Poisson point process on the positive line with intensity proportional to $t \, dt$.

1.4.3. *A statistical application.* Is it possible with probability greater than $\frac{1}{2}$ to tell the difference between two preferential attachment graphs started from nonisomorphic seeds and run for a

long time? This question is posed in the $\ell = 1$ case by Bubeck *et al.* [7], where the answer was determined to be yes as long as the degree sequences of the seeds are different. The crucial step is to separate the two graphs based on the maximum degree, which relies on a careful understanding of the maximum degree along the lines of Corollary 1. In fact, by exploiting the connection to the Brownian CRT just mentioned (along with other results), the answer to the question can be strengthened to yes as long as the two seed graphs are nonisomorphic; see [9].

1.4.4. *Organization of the paper.* To show Theorems 1 and 2, we relate the weight distributions of both models N_ℓ and L_ℓ to a single infinite-color urn model that generalizes the single color models considered in [13], [19], and [21]; urn models frequently appear when studying preferential attachment, see, for example, [3], [5], [19], [20], [22], and [25]. In the next section we define the relevant infinite-color urn model and make explicit the equivalence to the preferential attachment models under study. In Section 2 we state and prove a general approximation result from which Theorem 1 follows. Section 3 contains the proof of Theorem 2.

1.5. An infinite-color urn model

Consider the following infinite-color urn model. At step 0 there are $s \geq 1$ distinct colors present in the urn, and we assume that these colors are labelled from 1 to s . Fix an integer $\ell \geq 1$. At the n th step, a ball is picked at random from the urn and returned along with an additional ball of the same color. Additionally, if n is a multiple of ℓ , a ball of (the new) color $s + n/\ell$ is added after the n th draw.

We are interested in the cumulative color counts; that is, for each $k \geq 1$, let $M_k(n)$ be the number of balls of colors 1 to k in the urn after the n th draw (and possible immigration) has been completed. Let m_k be the number of balls of colours 1 to k at time 0, so that $M_k(0) = m_k$. In order to avoid degeneracies, we will assume that, at time 0, at least one ball of each color from 1 to s is present; that is, $m_1 \geq 1$ and $m_{k+1} - m_k \geq 1$ for $1 \leq k < s$.

The following result makes explicit the connection between the urn model just described and the preferential attachment models under study. It follows from straightforward considerations.

Lemma 1. *Fix a seed graph $G(0)$ with s vertices, and let d_1, \dots, d_s be the initial weight sequence. Let $\ell \geq 1$ and consider either the situation of*

- (i) *model N_ℓ for the graph sequence, and an infinite-color urn model having initially d_k balls of color k , where $1 \leq k \leq s$; or*
- (ii) *model L_ℓ for the graph sequence, and an infinite-color urn model having initially d_k balls of color k , where $1 \leq k \leq s$, and one ball of color $s + 1$.*

Then, for any r and $n > r$,

$$(d_1(n), d_1(n) + d_2(n), \dots, d_1(n) + \dots + d_r(n)) \stackrel{D}{=} (M_1(\ell n), \dots, M_r(\ell n)).$$

2. Proof of Theorem 1

Theorem 1 easily follows from the next result (proved immediately after its statement), the fact that linear transformations of convex sets are convex, and Lemma 1.

Proposition 2. *Fix $r > s$. Let B_1, \dots, B_{r-1} and Z_r be independent random variables such that*

$$B_k \sim \begin{cases} \text{beta}(m_k, m_{k+1} - m_k) & \text{if } 1 \leq k < s, \\ \text{beta}(m_s + (\ell + 1)(k - s) + \ell, 1) & \text{if } s \leq k < r, \end{cases}$$

and $Z_r \sim \text{GG}(m_s + (\ell + 1)(r - s) + \ell, \ell + 1)$. Define

$$Z_k = B_k \dots B_{r-1} Z_r, \quad 1 \leq k < r, \quad Z = (Z_1, \dots, Z_r),$$

and let

$$W(n) = (W_1(n), \dots, W_r(n)) = \frac{\ell^{\ell/(\ell+1)}}{(\ell + 1)n^{\ell/(\ell+1)}} (M_1(n), \dots, M_r(n)).$$

Then there is a positive constant $C = C(r, \ell, m_s)$, independent of n , such that

$$\sup_K |\mathbb{P}[W(n) \in K] - \mathbb{P}[Z \in K]| \leq \frac{C}{n^{\ell/(\ell+1)}} \quad \text{for all } n, \tag{1}$$

where the supremum ranges over all convex sets $K \subset \mathbb{R}^r$.

We need some intermediate lemmas to prove Proposition 2. Denote by $\mathcal{P}(\frac{b}{w}; m)$ the distribution of white balls in a classical Pólya urn after m completed draws, starting with b black and w white balls. Denote by $\mathcal{P}_{\text{Im}}^\ell(\frac{b}{w}; m)$ the number of white balls after m completed steps in the following Pólya urn with immigration, starting with b black and w white balls: at the n th step, a ball is picked at random from the urn and returned along with an additional ball of the same color; additionally, if n is a multiple of ℓ then a black ball is added after the n th draw and return.

Lemma 2. Let $p = k - s + 1$. If $k \geq s$, we have

$$M_k(n) \sim \mathcal{P}_{\text{Im}}^\ell \left(\frac{1}{m_s + \ell p + (k - s)}; n - \ell p \right). \tag{2}$$

Furthermore, conditionally on $M_{k+1}(n)$, we have

$$M_k(n) \sim \begin{cases} \mathcal{P} \left(\frac{m_{k+1} - m_k}{m_k}; M_{k+1}(n) - m_{k+1} \right) & \text{if } 1 \leq k < s, \\ \mathcal{P} \left(\frac{1}{m_s + \ell p + (k - s)}; M_{k+1}(n) - m_s - (\ell + 1)p \right) & \text{if } k \geq s. \end{cases} \tag{3}$$

Proof. To prove (2), first note that the number of balls having a color from the set $\{1, \dots, k\}$ is deterministic up to the point where the first ball of color $k + 1$ appears in the urn; this is the case after $\ell p = \ell(k - s + 1)$ completed draws. At that time we have $M_k(\ell p) + 1 = m_s + (\ell + 1)p$, so that $M_k(\ell p) = m_s + \ell p + (k - s)$. After that, consider all balls of colors $\{1, \dots, k\}$ as ‘white’ balls and all balls of colors $\{k + 1, \dots\}$ as ‘black’ balls. The number of white balls for the remaining $n - \ell p$ steps now behaves exactly like $\mathcal{P}_{\text{Im}}^\ell(\frac{b}{w}; m)$ with $b = 1, w = m_s + \ell p + (k - s)$, and $m = n - \ell p$.

To prove the second line of (3), consider all balls of colors $\{1, \dots, k\}$ as white balls and balls of color $k + 1$ as black balls. After time ℓp the number of white balls now behaves exactly like $\mathcal{P}(\frac{b}{w}; m)$ with $b = 1, w = M_k(\ell p)$, and m being the number of times a ball among colors $\{1, \dots, k + 1\}$ was picked after time p , which is just $M_{k+1}(n) - m_s - (\ell + 1)p$.

The argument to prove the first line of (3) is similar and, therefore, omitted. □

We will need the following coupling of Pólya urns and beta variables that is a generalization (from the $b = 1$ case) of [21, Lemma 4.4]; for related distributional approximation results, see [11].

Lemma 3. Fix positive integers $b, w,$ and $n.$ There is a coupling (X, Y) with $X \sim \mathcal{P}(\frac{b}{w}; n)$ and $Y \sim \text{beta}(w, b),$ such that almost surely,

$$|X - nY| < \frac{b(4w + b + 1)}{2}. \tag{4}$$

Proof. We use induction over $b,$ and start with the base case $b = 1.$ Let V_0, \dots, V_{w-1} be independent and uniformly distributed on the interval $[0, 1].$ By a well known representation of the distribution $\text{beta}(w, 1),$ we can choose

$$Y := \max(V_0, \dots, V_{w-1}).$$

To construct $X,$ first note that for $\mathcal{P}(\frac{1}{w}; m)\{A\}$ denoting the probability the relevant Pólya urn law puts on the set $A \subset \mathbb{Z},$ we have

$$\mathcal{P}\left(\frac{1}{w}; m\right)\{w, \dots, t\} = \prod_{k=0}^{w-1} \frac{t - k}{m + w - k}$$

for all $m \geq 0$ and for $w \leq t \leq w + m$ (see, for example, [10, Equation (2.4), p. 121]). For each $m \geq 0,$ let

$$N(m) := \max_{0 \leq k \leq w-1} (k + \lceil (m + w - k)V_k \rceil).$$

It is easy to see that the cumulative distribution function of $N(m)$ is that of $\mathcal{P}(\frac{1}{w}; m)$ for each $m,$ and that

$$|N(m) - mY| \leq w + 1 \quad \text{for all } m \geq 0. \tag{5}$$

Letting $X := N(n),$ (4) follows for the $b = 1$ case. As a side remark, note that, although $N(m) \sim \mathcal{P}(\frac{1}{w}; m)$ for each $m,$ the joint distribution of $(N(0), N(1), \dots)$ is not that of a Pólya urn process!

To prove the inductive step, assume we have constructed $N_{b-1}(0), N_{b-1}(1), \dots$ and Y_{b-1} such that $N_{b-1}(m) \sim \mathcal{P}(\frac{b-1}{w}; m)$ for all $m \geq 0, Y_{b-1} \sim \text{beta}(w, b - 1),$ and

$$|N_{b-1}(m) - mY_{b-1}| \leq \frac{(b - 1)(4w + b)}{2} \quad \text{for all } m \geq 0.$$

Now, let $Y'_b \sim \text{beta}(w, 1)$ be independent of all else and $N'(0), N'(1), \dots$ be defined and coupled to Y' as in the base case, that is, $N'(m) \sim \mathcal{P}(\frac{1}{w}; m)$ such that $|N'(m) - mY'| \leq w + 1.$ Define

$$N_b(m) := N'(N_{b-1}(m) - (w + b - 1)).$$

It is not difficult to see that $N_b(m) \sim \mathcal{P}(\frac{b}{w}; m).$ Also, it is not difficult to see that $Y_b := Y_{b-1}Y' \sim \text{beta}(w, b).$ Noting from (5) that, for any $y > 0,$

$$|N'(m) - yY'| \leq |N'(m) - mY'| + |m - y|Y' \leq (w + 1) + |m - y|,$$

we have

$$\begin{aligned} |N_b(m) - mY_b| &= |N'(N_{b-1}(m) - (w + b - 1)) - mY_{b-1}Y'| \\ &\leq (w + 1) + |N_{b-1}(m) - (w + b - 1) - mY_{b-1}| \\ &\leq (w + 1) + (w + b - 1) + \frac{(b - 1)(4w + b)}{2} \\ &= \frac{b(4w + b + 1)}{2}. \end{aligned}$$

This concludes the inductive step, where (4) is just the $m = n$ case. □

The last ingredient we need to prove Proposition 2 is some moment information.

Lemma 4. For any $k > s$, and $q = 1, 2, \dots$, we have

$$\mathbb{E}M_k(n)^q \simeq n^{q\ell/(\ell+1)}, \tag{6}$$

and, for $w = m_s + (\ell + 1)(k - s) + \ell$ and $t = n - \ell(k - s + 1)$,

$$\begin{aligned} & \mathbb{E}\{M_k(n)(M_k(n) + 1) \cdots (M_k(n) + \ell)\} \\ &= w \frac{(w + 1 + \lfloor (t - 1)/\ell \rfloor + t) \cdots (w + 1 + \lfloor (t - 1)/\ell \rfloor + t + \ell)}{w + 1 + (\ell + 1)\lfloor (t - 1)/\ell \rfloor + \ell} \\ &= (m_s + (\ell + 1)(k - s) + \ell)n^\ell \left(\frac{\ell + 1}{\ell}\right)^\ell (1 + O(n^{-1})). \end{aligned} \tag{7}$$

Furthermore,

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left\{ \frac{n^{\ell/(\ell+1)}}{M_k(n)} \right\} < \infty. \tag{8}$$

Proof. The asymptotic (6) follows from [21, Lemma 4.1]. From that lemma, we also have, for $Y \sim \mathcal{P}_{\text{Im}}^\ell(w; t)$,

$$\mathbb{E}\{Y(Y + 1) \cdots (Y + \ell)\} = \prod_{j=0}^{\ell} (w + j) \prod_{i=0}^{t-1} \left(1 + \frac{\ell + 1}{w + 1 + i + \lfloor i/\ell \rfloor}\right).$$

Setting $T = \lfloor (t - 1)/\ell \rfloor$, using careful bookkeeping and a telescoping argument, we obtain

$$\begin{aligned} & \mathbb{E}\{Y(Y + 1) \cdots (Y + \ell)\} \\ &= \prod_{j=0}^{\ell} (w + j) \prod_{i=0}^{t-1} \left(1 + \frac{\ell + 1}{w + 1 + i + \lfloor i/\ell \rfloor}\right) \\ &= \prod_{j=0}^{\ell} (w + j) \prod_{k=0}^{T-1} \prod_{i=0}^{\ell-1} \left(1 + \frac{\ell + 1}{w + 1 + i + k(\ell + 1)}\right) \prod_{m=\ell T}^{t-1} \left(1 + \frac{\ell + 1}{w + 1 + m + T}\right) \\ &= w \prod_{i=0}^{\ell-1} (w + 1 + i + (\ell + 1)T) \prod_{m=\ell T}^{t-1} \left(1 + \frac{\ell + 1}{w + 1 + m + T}\right) \\ &= w \prod_{i=0}^{\ell-1} (w + 1 + i + (\ell + 1)T) \prod_{m=0}^{t-1-\ell T} \left(1 + \frac{\ell + 1}{w + 1 + m + (\ell + 1)T}\right) \\ &= w \prod_{i=t-\ell T}^{\ell-1} (w + 1 + i + (\ell + 1)T) \prod_{m=0}^{t-1-\ell T} (w + 1 + m + (\ell + 1)(T + 1)) \\ &= w \prod_{i=0}^{\ell+\ell T-t-1} (w + 1 + i + T + t) \prod_{m=\ell+\ell T-t+1}^{\ell} (w + 1 + m + T + t) \\ &= w \frac{(w + 1 + T + t) \cdots (w + 1 + T + t + \ell)}{w + 1 + (\ell + 1)T + \ell}. \end{aligned}$$

Setting $w = m_s + (\ell + 1)(k - s) + \ell$ and $t = n - \ell(k - s + 1)$ as per (2) yields (7).

In order to prove (8), let

$$X \sim \mathcal{P}_{\text{Im}}^\ell \left(\begin{matrix} 1 \\ w+1 \end{matrix}; t \right) \quad \text{and} \quad Y \sim \mathcal{P}_{\text{Im}}^\ell \left(\begin{matrix} 1 \\ w \end{matrix}; t \right),$$

where $w = m_s + (\ell + 1)(k - s) + \ell - 1$. From (2) and [21, Lemma 4.2], we have

$$\mathbb{E}f(X) = \frac{\mathbb{E}\{Yf(Y+1)\}}{\mathbb{E}Y}$$

for any bounded function f ; in particular, for the function $f(x) = 1/x$, bounded when $x \geq 1$, we have

$$\mathbb{E}X^{-1} = \mathbb{E} \left\{ \frac{Y}{(Y+1)\mathbb{E}Y} \right\} \leq \frac{1}{\mathbb{E}Y}.$$

By (6), $\mathbb{E}Y \asymp n^{\ell/(\ell+1)}$, from which (8) easily follows. □

Note that the Landau- O notation of the lemma contains a constant that depends on ℓ, k , and both of m_s, s . But here and below we ignore the dependence of constants on s when also depending on m_s through the inequalities $1 \leq s \leq m_s$.

Proof of Proposition 2. To ease notation, we fix n and drop it in our variable notation and write, for example, W_k instead of $W_k(n)$. For $k < l$, let

$$K_{k,l} = B_k \dots B_l, \quad V_{k,l} = K_{k,l-1}W_l, \quad V_{l,l} = W_l.$$

We may assume that B_1, \dots, B_{r-1} are all independent of $W = W(n)$. Fix a convex subset $K \subset \mathbb{R}^r$, assuming without loss of generality that K is closed, and let h be the indicator function of K . We proceed in two major steps by writing

$$\begin{aligned} \mathbb{P}[W \in K] - \mathbb{P}[Z \in K] &= \mathbb{E}h(W) - \mathbb{E}h(Z) \\ &= \mathbb{E}\{h(W) - h(V_{1,r}, \dots, V_{r,r})\} + \mathbb{E}\{h(V_{1,r}, \dots, V_{r,r}) - h(Z)\} \\ &=: \mathbb{E}R_1 + \mathbb{E}R_2. \end{aligned}$$

In order to bound R_1 , first write it as the telescoping sum

$$\begin{aligned} R_1 &= h(V_{1,1}, \dots, V_{r,r}) - h(V_{1,r}, V_{2,r}, \dots, V_{r,r}) \\ &= \sum_{k=1}^{r-1} [h(V_{1,k}, \dots, V_{k-1,k}, V_{k,k}, V_{k+1,k+1}, \dots, V_{r,r}) \\ &\quad - h(V_{1,k+1}, \dots, V_{k-1,k+1}, V_{k,k+1}, V_{k+1,k+1}, \dots, V_{r,r})] \\ &=: \sum_{k=1}^{r-1} R_{1,k}. \end{aligned}$$

Fix k , and define b_k and w_k to be the parameters of the Pólya urn in Lemma 2, that is,

$$b_k = \begin{cases} m_{k+1} - m_k & \text{if } 1 \leq k < s, \\ 1 & \text{if } k \geq s, \end{cases}$$

$$w_k = \begin{cases} m_k & \text{if } 1 \leq k < s, \\ m_s + \ell(k - s + 1) + (k - s) & \text{if } k \geq s. \end{cases}$$

Conditioning on M_{k+1} , we can use Lemmas 2 and 3 to conclude that there is a coupling (X_k, Y_k) with

$$X_k \sim \mathcal{P}\left(\begin{matrix} b_k \\ w_k \end{matrix}; M_{k+1} - c_k\right), \quad Y_k \sim \text{beta}(w_k, b_k),$$

such that almost surely,

$$|X_k - (M_{k+1} - c_k)Y_k| \leq \frac{b_k(4w_k + b_k + 1)}{2},$$

where

$$c_k = \begin{cases} m_{k+1} & \text{if } 1 \leq k < s, \\ m_s + (\ell + 1)(k - s + 1) & \text{if } k \geq s. \end{cases}$$

Hence, there is a constant $C(k, \ell, m_s)$ and a random variable E_k with $|E_k| \leq C(k, \ell, m_s)$ almost surely, such that

$$\frac{X_k}{M_{k+1}} = Y_k + \frac{E_k}{M_{k+1}}.$$

Define

$$V'_{j,k} := \frac{\ell^{\ell/(\ell+1)}}{(\ell + 1)n^{\ell/(\ell+1)}} K_{j,k-1} X_k = K_{j,k-1} \frac{X_k}{M_{k+1}} W_{k+1} \quad \text{for } 1 \leq j < k,$$

$$V''_{j,k+1} := \frac{\ell^{\ell/(\ell+1)}}{(\ell + 1)n^{\ell/(\ell+1)}} K_{j,k-1} Y_k M_{k+1} = K_{j,k-1} Y_k W_{k+1} \quad \text{for } 1 \leq j < k + 1,$$

and

$$V'_{j,j} := V_{j,j} \quad \text{for } j \geq k, \quad V''_{j,j} := V_{j,j} \quad \text{for } j \geq k + 1.$$

Note that by Lemma 2, $\mathcal{L}(V'_{j,k}) = \mathcal{L}(V_{j,k})$ and $\mathcal{L}(V''_{j,k+1}) = \mathcal{L}(V_{j,k+1})$, hence, we have

$$\begin{aligned} R_{1,k} &\stackrel{D}{=} h(V'_{1,k}, \dots, V'_{k-1,k}, V'_{k,k}, V'_{k+1,k+1}, \dots, V'_{r,r}) \\ &\quad - h(V''_{1,k+1}, \dots, V''_{k-1,k+1}, V''_{k,k+1}, V''_{k+1,k+1}, \dots, V''_{r,r}) \\ &= g\left(Y_k + \frac{E_k}{M_{k+1}}\right) - g(Y_k), \end{aligned}$$

where we define (in notation anticipating future conditioning)

$$g(x) := h(K_{1,k-1}xW_{k+1}, \dots, K_{k-1,k-1}xW_{k+1}, W_{k+1}, \dots, W_r).$$

Let \mathcal{F}_k be the σ -algebra generated by B_1, \dots, B_{k-1}, E_k , and W (hence, also $(M_j)_{j=1}^r$). Since h is the indicator function of a convex set, and since linear transformations of convex sets result again in convex sets, conditional on \mathcal{F}_k , it follows that g is of the form $g(x) = \mathbf{1}(a \leq x \leq b)$ for $-\infty \leq a \leq b \leq \infty$ which are random but \mathcal{F}_k -measurable. Hence,

$$\begin{aligned} \mathbb{E}\{\mathbf{1}\{R_{1,k}\} \mid \mathcal{F}_k\} &\leq \mathbb{P}\left[a - \frac{|E_k|}{M_{k+1}} \leq Y_k \leq a + \frac{|E_k|}{M_{k+1}} \mid \mathcal{F}_k\right] \\ &\quad + \mathbb{P}\left[b - \frac{|E_k|}{M_{k+1}} \leq Y_k \leq b + \frac{|E_k|}{M_{k+1}} \mid \mathcal{F}_k\right] \\ &\leq 2 \sup_{x \in \mathbb{R}} \mathbb{P}\left[x - \frac{C}{M_{k+1}} \leq Y_k \leq x + \frac{C}{M_{k+1}} \mid \mathcal{F}_k\right] \\ &\leq \frac{C'}{M_{k+1}} \end{aligned}$$

for some constant $C' = C'(k, \ell, m_s)$, and where the last inequality follows from basic properties of the beta distribution. Hence, by (8) of Lemma 4,

$$\mathbb{E}|R_{1,k}| \leq \mathbb{E} \left\{ \frac{C'}{M_{k+1}} \right\} \leq \frac{C''}{n^{\ell/(\ell+1)}} \tag{9}$$

for some constant $C'' = C''(k, \ell, m_s)$.

In order to bound R_2 , write

$$\begin{aligned} R_2 &= h((K_{1,r-1}, \dots, K_{r-1,r-1}, 1)W_r) - h((K_{1,r-1}, \dots, K_{r-1,r-1}, 1)Z_r) \\ &= g(W_r) - g(Z_r), \end{aligned}$$

where now

$$g(x) = h((K_{1,r-1}, \dots, K_{r-1,r-1}, 1)x).$$

Given $K_{1,r-1}, \dots, K_{r-1,r-1}$, the function g is again of the same form as in the first part of the proof, so that

$$\mathbb{E}|R_2| \leq 2 d_{\text{Kol}}(\mathcal{L}(W_r), \mathcal{L}(Z_r)), \tag{10}$$

where d_{Kol} denotes the Kolmogorov distance, i.e. the supremum distance between distribution functions. Define the scaling constants

$$\tilde{\mu}_n = \frac{(\ell + 1)n^{\ell/(\ell+1)}}{\ell^{\ell/(\ell+1)}} \quad \text{and} \quad \mu_n = \left(\frac{(\ell + 1)\mathbb{E}M_r(n)^{\ell+1}}{m_s + (\ell + 1)(r - s) + \ell} \right)^{1/(\ell+1)}.$$

Since the Kolmogorov distance is scale invariant,

$$\begin{aligned} d_{\text{Kol}}(\mathcal{L}(W_r), \mathcal{L}(Z_r)) &= d_{\text{Kol}}\left(\mathcal{L}\left(\frac{M_r(n)}{\mu_n}\right), \mathcal{L}\left(\frac{Z_r \tilde{\mu}_n}{\mu_n}\right)\right) \\ &\leq d_{\text{Kol}}\left(\mathcal{L}\left(\frac{M_r(n)}{\mu_n}\right), \mathcal{L}(Z_r)\right) + d_{\text{Kol}}\left(\mathcal{L}(Z_r), \mathcal{L}\left(\frac{Z_r \tilde{\mu}_n}{\mu_n}\right)\right) \\ &:= R_{2,1} + R_{2,2}. \end{aligned} \tag{11}$$

From [21, Theorem 1.2], we have $R_{2,1} = O(n^{-\ell/(\ell+1)})$ and, noting that the density of Z_r is bounded, standard arguments yield $R_{2,2} = O(|1 - \tilde{\mu}_n/\mu_n|)$. To bound $|1 - \tilde{\mu}_n/\mu_n|$, use (6) and (7) to find

$$\begin{aligned} \mathbb{E}M_r(n)^{\ell+1} &= \mathbb{E}\{M_r(n) \cdots (M_r(n) + \ell)\} + O(n^{\ell^2/(\ell+1)}) \\ &= (m_s + (\ell + 1)(r - s) + \ell)n^\ell \left(\frac{\ell + 1}{\ell}\right)^\ell (1 + O(n^{-1})) + O(n^{\ell^2/(\ell+1)}) \\ &= (m_s + (\ell + 1)(r - s) + \ell)n^\ell \left(\frac{\ell + 1}{\ell}\right)^\ell + O(n^{\ell^2/(\ell+1)}). \end{aligned}$$

Using this last expression, we easily find

$$\frac{\mu_n}{\tilde{\mu}_n} = (1 + O(n^{-\ell/(\ell+1)}))^{1/(\ell+1)}.$$

Now, using the fact that, for $0 < x < 1$, $(1 + x)^{1/(\ell+1)} - 1 \leq x/(\ell + 1)$, we find $\mu_n/\tilde{\mu}_n = 1 + O(n^{-\ell/(\ell+1)})$ and so $R_{2,2} = O(n^{-\ell/(\ell+1)})$. Now collecting the bounds (9)–(11) we are able to prove (1). □

3. Proof of Theorem 2

Since Theorem 1 establishes convergence of finite-dimensional distributions, to prove the claimed convergence of $(\tilde{D}(n))_{n \geq 1}$, we need to show tightness in l_p for the relevant p . In the next section we state and prove an l_p tightness criteria in terms of moment conditions and then in Section 3.2 we bound the relevant moments appropriately. In Section 3.3 we establish the continuity of the ordering function on l_p and put everything together to prove Theorem 2 in Section 3.4.

3.1. Tightness in l_p

The following result is essentially Suquet’s [27, Section 6, Theorem 16], but we give a self-contained proof.

Lemma 5. *Let $X(n) = (X_1(n), X_2(n), \dots)$, $n \geq 1$, be a sequence of l_p -valued random elements, where $1 \leq p < \infty$. If*

- (i) $\sup_{n \geq 1} \sum_{i \geq 1} \mathbb{E}|X_i(n)|^p < \infty$, and
- (ii) $\lim_{k \rightarrow \infty} \sup_{n \geq 1} \sum_{i \geq k} \mathbb{E}|X_i(n)|^p = 0$,

then the sequence of probability measures $(\mathbb{P}[X(n) \in \cdot])_{n \geq 1}$ is tight in the l_p -topology.

Proof. By the Fréchet–Kolmogorov theorem, $B \subset l_p$ is relatively compact if it is bounded and

$$\lim_{k \rightarrow \infty} \sup_{x \in B} \sum_{i \geq k} |x_i|^p = 0.$$

Thus, for any $C > 0$ and any strictly increasing sequence of numbers $1 \leq k_1 < k_2 < \dots$, the set

$$\mathcal{K}(C, (k_m)_{m \geq 1}) := \left\{ x \in l_p : \sum_{i \geq 1} |x_i|^p \leq C, \text{ and } \sum_{i \geq k_m} |x_i|^p \leq \frac{1}{m} \text{ for all } m \geq 1 \right\}$$

is l_p -compact.

Now, fix $\varepsilon > 0$. From (i) we conclude that there is C such that

$$\mathbb{P}\left[\sum_{i \geq 1} |X_i(n)|^p > C\right] \leq \frac{1}{C} \sum_{i \geq 1} \mathbb{E}|X_i(n)|^p \leq \frac{\varepsilon}{2} \quad \text{for all } n \geq 1.$$

From (ii) we conclude that there is a sequence $1 \leq k_1 < k_2 < \dots$ such that, for all $m \geq 1$,

$$\mathbb{P}\left[\sum_{i \geq k_m} |X_i(n)|^p > \frac{1}{m}\right] \leq m \sum_{i \geq k_m} \mathbb{E}|X_i(n)|^p \leq \frac{\varepsilon}{2^{m+1}} \quad \text{for all } n \geq 1.$$

For this C and $(k_m)_{m \geq 1}$, we conclude that, for all $n \geq 1$,

$$\begin{aligned} \mathbb{P}[X(n) \notin \mathcal{K}(C, (k_m)_{m \geq 1})] &\leq \mathbb{P}\left[\sum_{i \geq 1} |X_i(n)|^p > C\right] + \sum_{m \geq 1} \mathbb{P}\left[\sum_{i \geq k_m} |X_i(n)|^p > \frac{1}{m}\right] \\ &\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \sum_{m \geq 1} \frac{1}{2^m} \\ &= \varepsilon. \end{aligned} \quad \square$$

3.2. Moment bounds

The following moment bounds are used in order to apply Lemma 5.

Lemma 6. *Let $\ell \geq 1$, and let $D_k(n)$ be as in Theorem 1. There is a constant $C = C(\ell, m_s, s)$ such that*

$$\mathbb{E}D_k(n)^{\ell+1} \leq Ck^{-\ell}.$$

Moreover, in the special cases N_1 and L_1 ,

$$\mathbb{E}D_k(n)^4 \leq Ck^{-2}.$$

Proof. By Lemma 1, it is enough to prove the theorem for the appropriate equivalent urn model (possibly adjusting s and m_s , each by at most one). Recall the definition of $M_k(n)$ as in Proposition 2, set $U_1(n) = M_1(n)$, and, for $k = 2, 3, \dots$, let $U_k(n) = M_k(n) - M_{k-1}(n)$, where we set $U_k(n) = 0$ if k is a color that has not been added by time n . Note that $\mathcal{L}(U_k(n)) = \mathcal{L}(D_k(n))$.

First note that the statement of the theorem is trivial if $U_k(n) = 0$, that is, if color k has not yet appeared; in particular, if $k - s > n + 1$. Now, for $k > s$,

$$\mathcal{L}(U_k(n) \mid M_k(n)) = \mathcal{P}\left(m_s + (k - s)(\ell + 1) - 1; M_k(n) - m_s - (k - s)(\ell + 1)\right).$$

Let the random variable $B \sim \text{beta}[1, m_s + (k - s)(\ell + 1) - 1]$ be independent of $M_k(n)$. Conditional on B and $M_k(n)$, let $X(M_k(n), B)$ be binomial with parameters $M_k(n) - m_s - (k - s)(\ell + 1)$ and B . By the de Finetti representation of the classical Pólya urn, we have

$$\mathcal{L}(U_k(n) \mid M_k(n)) = \mathcal{L}((1 + X(M_k(n), B)) \mid M_k(n)). \tag{12}$$

Hölder’s inequality implies that, for nonnegative x, y ,

$$(x + y)^p \leq 2^{p-1}(x^p + y^p),$$

and so starting from (12), we have

$$\mathbb{E}U_k(n)^p \leq 2^{p-1}(1 + \mathbb{E}X(M_k(n), B)^p). \tag{13}$$

Now, note that, if $\mathcal{L}(Y) = \text{bin}(N, q)$ then for positive integer p , and denoting Stirling numbers of the second kind by $\left\{ \begin{smallmatrix} p \\ j \end{smallmatrix} \right\}$ (and note these are nonnegative),

$$\mathbb{E}Y^p = \sum_{j=0}^p \left\{ \begin{smallmatrix} p \\ j \end{smallmatrix} \right\} \mathbb{E}\{Y(Y - 1) \cdots (Y - j + 1)\} \leq \sum_{j=0}^p \left\{ \begin{smallmatrix} p \\ j \end{smallmatrix} \right\} (Nq)^j.$$

So from (13), condition on $M_k(n)$ and B to find

$$\mathbb{E}U_k(n)^p \leq 2^{p-1} \left(1 + \sum_{j=0}^p \left\{ \begin{smallmatrix} p \\ j \end{smallmatrix} \right\} \mathbb{E}M_k(n)^j \mathbb{E}B^j \right). \tag{14}$$

Standard formulae for beta moments imply that

$$\mathbb{E}B^j = \frac{\Gamma(j + 1)\Gamma(m_s + (k - s)(\ell + 1))}{\Gamma(m_s + (k - s)(\ell + 1) + j)} \leq Ck^{-j}, \tag{15}$$

where $C = C(\ell, m_s, s)$ is a constant. By Jensen's (or Hölder's) inequality, for $j \leq p$,

$$\mathbb{E}M_k(n)^j \leq (\mathbb{E}M_k(n)^p)^{j/p}. \tag{16}$$

Set $p = \ell + 1$, and use the fact that $M_k(n)^{\ell+1} \leq M_k(n)(M_k(n) + 1) \cdots (M_k(n) + \ell)$. From (7),

$$\begin{aligned} & \mathbb{E}\{M_k(n)(M_k(n) + 1) \cdots (M_k(n) + \ell)\} \\ &= (m_s + (\ell + 1)(k - s) + \ell) \\ & \times \frac{\prod_{j=0}^{\ell} (m_s + k - s + 1 + \lfloor (n - \ell(k - s + 1) - 1)/\ell \rfloor + n + j)}{m_s + (\ell + 1)(k - s + 1) + (\ell + 1)\lfloor (n - \ell(k - s + 1) - 1)/\ell \rfloor + \ell} \\ & \leq (m_s + (\ell + 1)k + \ell) \frac{(m_s + k - s + 1 + (n - \ell(k - s + 1) - 1)/\ell + n + \ell)^{\ell+1}}{m_s + (\ell + 1)(k - s + 1) + (\ell + 1)(n - \ell(k - s + 1) - 1)/\ell - 1} \\ & \leq (m_s + (\ell + 1)k + \ell) \frac{(m_s + n((\ell + 1)/\ell) + \ell)^{\ell+1}}{m_s + n((\ell + 1)/\ell) - (2\ell + 1)/\ell} \\ & \leq C'kn^{\ell}, \end{aligned}$$

where $C' = C'(\ell, m_s)$ is a constant. These moment estimates together with (14) and (15) imply that

$$\mathbb{E}U_k(n)^{\ell+1} \leq C'' \sum_{j=0}^{\ell+1} (kn^{\ell})^{j/(\ell+1)} k^{-j} = C'' \sum_{j=0}^{\ell+1} \left(\frac{n}{k}\right)^{j\ell/(\ell+1)} \leq C''' \left(\frac{n}{k}\right)^{\ell}, \tag{17}$$

where C'' and C''' are constants depending only on ℓ, m_s, s and the last inequality is because we may assume that $n \geq k - s - 1$. Scaling $U_k(n)$ by $n^{-\ell/(\ell+1)}$, we have

$$\mathbb{E}\{(n^{-\ell/(\ell+1)}U_k(n))^{\ell+1}\} \leq C'''k^{-\ell},$$

which proves the first assertion of the lemma. For the second assertion, follow the proof up to (16) and now set $\ell = 1$ and $p = 4$. Again $\mathbb{E}M_k(n)^4$ can be bounded by the fourth rising factorial moment and now [21, Lemma 4.1] implies that, for $w = m_s + 2(k - s) + 1$ and $t = n - (k - s + 1)$,

$$\begin{aligned} & \mathbb{E}\{M_n(k) \cdots (M_n(k) + 3)\} \\ &= \prod_{j=0}^3 (w + j) \prod_{i=0}^{t-1} \frac{w + 1 + 2i + 4}{w + 1 + 2i} \\ &= w(w + 2)(w + 1 + 2t)(w + 1 + 2t + 2) \\ & \leq (w + 2)^2(w + 2t + 3) \\ &= (m_s + 2(k - s) + 3)^2(m_s + 2(k - s) + 1 + 2(n - (k - s) - 1) + 3)^2 \\ & \leq ck^2n^2, \end{aligned}$$

where $c = c(m_s, s)$ is a constant. Now applying the same argument as (17), it follows that, for some constants c' and c'' depending only on m_s and s ,

$$\mathbb{E}U_k(n)^4 \leq c' \sum_{j=0}^4 \mathbb{E}\{(n^2k^2)^{j/4}\}k^{-j} \leq c'' \left(\frac{n}{k}\right)^2.$$

Scaling $U_k(n)$ in this last equation yields the second assertion. □

3.3. Continuity of the ordering function

For $x \in l_p^+$, define the ordering function $x^\downarrow \in l_p^+$ as follows. Since $\lim_{i \rightarrow \infty} x_i = 0$, the sequence has a maximum. Let y_1 be the value of that maximum, remove it from the sequence x , and repeat, assigning consecutive values to y_2, y_3 , and so forth. Then, set $x^\downarrow := y$. Note that the ordering function is not just a permutation of the coordinates: $(\frac{1}{2}, 0, \frac{1}{4}, 0, \frac{1}{8}, 0, \dots)^\downarrow = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots)$. To see that $x^\downarrow \in l_p^+$, note that $\|x^\downarrow\|_p = \|x\|_p$, since every positive value appearing in x will appear in x^\downarrow , and zero values in x do not contribute to the norm.

Lemma 7. *For any $1 \leq p < \infty$, the ordering function is continuous in l_p^+ .*

Remark 3. As pointed out to us by a referee, the ordering function is in fact 1-Lipschitz in l_p ; this follows from [15, Theorem 3.5] applied to step functions. We present the lemma and proof for the sake of completeness.

Proof of Lemma 7. Since l_p^+ is a metric space, it is enough to show that the ordering function is sequentially continuous. Assume that $x(n) \rightarrow x$ in l_p^+ ; that is, $\|x(n) - x\|_p \rightarrow 0$ as $n \rightarrow \infty$. For $\alpha > 0$ (to be chosen later), define the set of indices

$$M_\alpha = M_\alpha(x) = \{i : x_i > \alpha\};$$

since $\lim_{i \rightarrow \infty} x_i = 0$, the set M_α is finite for any positive α . Define the sequences x' and x'' by $x'_i = x_i \mathbf{1}(i \in M_\alpha)$ and $x'' = x - x'$. Moreover, for each n , define the two sequences $x'(n)$ and $x''(n)$ by $x'_i(n) = x_i(n) \mathbf{1}(i \in M_\alpha)$ and $x''(n) = x(n) - x'(n)$ (note that the sequence $x(n)$ is decomposed with respect to M_α , which depends on x only).

Now, fix $\varepsilon > 0$, and choose α such that

$$\|x''\|_p \leq \frac{1}{4}\varepsilon,$$

and it is, moreover, possible to choose α such that

$$\max_{i \notin M_\alpha} x_i < \alpha < \min_{i \in M_\alpha} x_i. \tag{18}$$

Clearly, the first $|M_\alpha|$ elements of x^\downarrow are just the ordered nonzero elements of x' . Now, choose N large enough so that

- (i) $\|x - x(n)\|_p \leq \frac{1}{4}\varepsilon$ for all $n \geq N$,
- (ii) $x_i(n) > \alpha$ for all $i \in M_\alpha$ all $n \geq N$,
- (iii) $x_i(n) < \alpha$ for all $i \notin M_\alpha$ and all $n \geq N$, and
- (iv) $\|(x')^\downarrow - (x'(n))^\downarrow\|_p \leq \frac{1}{4}\varepsilon$ for all $n \geq N$.

Condition (i) can be achieved because $x(n) \rightarrow x$; conditions (ii) and (iii) can be achieved because $x(n) \rightarrow x$ implies uniform coordinatewise convergence and because of (18); condition (iv) can be achieved again by uniform coordinatewise convergence and the fact that the order statistics of finitely many values can be approximated to any arbitrary precision. It follows that, for any $n \geq N$,

$$\|x^\downarrow - (x(n))^\downarrow\|_p \leq \|(x')^\downarrow - (x'(n))^\downarrow\|_p + \|x''\|_p + \|x''(n)\|_p \leq \frac{1}{4}\varepsilon + \frac{1}{4}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon. \quad \square$$

3.4. Proof of Theorem 2

Proof. We first argue that almost surely $Y \in l_p^+$ for $p > \ell/(\ell + 1)$. First note that $Y_i = Z_i - Z_{i-1}$ and, according to Theorem 1, for $i > s$, $Z_i = \text{GG}(a_s + (\ell + 1)(i - s))$ and $Z_{i-1} = B_{i-1}Z_i$, where $B_{i-1} \sim \text{beta}(a_s + (\ell + 1)(i - s - 1), 1)$ and is independent of Z_i . Thus, we can write $Y_i = Z_i(1 - B_{i-1})$ and note that $(1 - B_{i-1}) \sim \text{beta}(1, a_s + (\ell + 1)(i - s - 1))$. From standard moment formula for gamma and beta distributions, we have

$$\begin{aligned} \mathbb{E}Y_i^p &= \frac{\Gamma(a_s/(\ell + 1) + i - s + p/(\ell + 1))}{\Gamma(a_s/(\ell + 1) + i - s)} \frac{\Gamma(1 + p)\Gamma(a_s + (\ell + 1)(i - s - 1) + 1)}{\Gamma(a_s + (\ell + 1)(i - s - 1) + 1 + p)} \\ &\leq Ci^{-p\ell/(\ell+1)}, \end{aligned}$$

where $C = C(p, \ell, a_s, s)$ is a constant and the inequality follows from $x^t\Gamma(x)/\Gamma(x + t) \rightarrow 1$ as $x \rightarrow \infty$. Clearly, we can extend the inequality to all $i \geq 1$. Therefore,

$$\mathbb{E} \sum_{i=1}^{\infty} Y_i^p = \sum_{i=1}^{\infty} \mathbb{E}Y_i^p \leq C \sum_{i=1}^{\infty} i^{-p\ell/(\ell+1)},$$

and this last term is finite for $p > (\ell + 1)/\ell$. Since the expectation of the sum is finite, the sum is almost surely finite.

The tightness of $(\tilde{D}(n))_{n \geq 1}$ follows by applying the moment bounds of Lemma 6 in Lemma 5 and then convergence to the claimed limit follows from the convergence of finite-dimensional distributions given by Theorem 1.

Finally, $(\tilde{D}(n)^\downarrow)_{n \geq 1}$ converges to \tilde{Y}^\downarrow because of the convergence of $(\tilde{D}(n))_{n \geq 1}$, the continuity of the order mapping given by Lemma 7. □

Acknowledgements

We thank Jim Pitman for the suggestion to study joint degree distributions for preferential attachment graphs and for pointers to the CRT literature, Sourav Chatterjee for the suggestion to look at the order statistics of the process, Rongfeng Sun for helpful discussions, and the anonymous referees for their valuable comments. E.P. thanks the Department of Mathematics and Statistics at the University of Melbourne for their hospitality during a visit when much of this work was completed, supported by a UofM ECR grant. A.R. was supported by NUS Research Grant R-155-000-124-112. N.R. was supported by ARC grant DP150101459.

References

- [1] ALDOUS, D. (1991). The continuum random tree. I. *Ann. Prob.* **19**, 1–28.
- [2] ALDOUS, D. (1993). The continuum random tree. III. *Ann. Prob.* **21**, 248–289.
- [3] ANTUNOVIĆ, T., MOSSEL, E. AND RÁČZ, M. Z. (2016). Coexistence in preferential attachment networks. *Combinatorics Prob. Comput.* **25**, 797–822.
- [4] BARABÁSI, A.-L. AND ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286**, 509–512.
- [5] BERGER, N., BORGS, C., CHAYES, J. T. AND SABERI, A. (2014). Asymptotic behavior and distributional limits of preferential attachment graphs. *Ann. Prob.* **42**, 1–40.
- [6] BOLLOBÁS, B., RIORDAN, O., SPENCER, J. AND TUSNÁDY, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18**, 279–290.
- [7] BUBECK, S., MOSSEL, E. AND RÁČZ, M. Z. (2015). On the influence of the seed graph in the preferential attachment model. *IEEE Trans. Network Sci. Eng.* **2**, 30–39.
- [8] COLLEVECCHIO, A., COTAR, C. AND LICALZI, M. (2013). On a preferential attachment and generalized Polya’s urn model. *Ann. Appl. Prob.* **23**, 1219–1253.
- [9] CURIEN, N., DUQUESNE, T., KORTCHEMSKI, I. AND MANOLESCU, I. (2015). Scaling limits and influence of the seed graph in preferential attachment trees. *J. Éc. Polytech. Math.* **2**, 1–34.

- [10] FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd edn. John Wiley, New York.
- [11] GOLDSTEIN, L. AND REINERT, G. (2013). Stein's method for the beta distribution and the Pólya–Eggenberger urn. *J. Appl. Prob.* **50**, 1187–1205.
- [12] JAMES, L. F. (2015). Generalized Mittag Leffler distributions arising as limits in preferential attachment models. Preprint. Available at <https://arxiv.org/abs/1509.07150v4>.
- [13] JANSON, S. (2006). Limit theorems for triangular urn schemes. *Prob. Theory Relat. Fields* **134**, 417–452.
- [14] KRAPIVSKY, P. L., REDNER, S. AND LEYVRAZ, F. (2000). Connectivity of growing random networks. *Phys. Rev. Lett.* **85**, 4629.
- [15] LIEB, E. H. AND LOSS, M. (2001). *Analysis* (Graduate Stud. Math. **14**), 2nd edn. American Mathematical Society, Providence, RI.
- [16] MÓRI, T. F. (2005). The maximum degree of the Barabási–Albert random tree. *Combinatorics Prob. Comput.* **14**, 339–348.
- [17] NEWMAN, M. E. J. (2003). The structure and function of complex networks. *SIAM Rev.* **45**, 167–256.
- [18] NEWMAN, M., BARABÁSI, A.-L. AND WATTS, D. J. (2006). *The Structure and Dynamics of Networks*. Princeton University Press.
- [19] PEKÖZ, E. A., RÖLLIN, A. AND ROSS, N. (2013). Degree asymptotics with rates for preferential attachment random graphs. *Ann. Appl. Prob.* **23**, 1188–1218.
- [20] PEKÖZ, E. A., RÖLLIN, A. AND ROSS, N. (2013). Total variation error bounds for geometric approximation. *Bernoulli* **19**, 610–632.
- [21] PEKÖZ, E. A., RÖLLIN, A. AND ROSS, N. (2016). Generalized gamma approximation with rates for urns, walks and trees. *Ann. Prob.* **44**, 1776–1816.
- [22] PEMANTLE, R. (2007). A survey of random processes with reinforcement. *Prob. Surveys* **4**, 1–79.
- [23] PITMAN, J. (1999). Brownian motion, bridge, excursion, and meander characterized by sampling at independent uniform times. *Electron. J. Prob.* **4**, 33pp.
- [24] PITMAN, J. (2006). *Combinatorial Stochastic Processes* (Lecture Notes Math. **1875**). Springer, Berlin.
- [25] ROSS, N. (2013). Power laws in preferential attachment graphs and Stein's method for the negative binomial distribution. *Adv. Appl. Prob.* **45**, 876–893.
- [26] RUDAS, A., TÓTH, B. AND VALKÓ, B. (2007). Random trees and general branching processes. *Random Structures Algorithms* **31**, 186–202.
- [27] SUQUET, C. (1999). Tightness in Schauder decomposable Banach spaces. In *Proc. St. Petersburg Mathematical Society* (Amer. Math. Soc. Transl. Ser. **193**), Vol. V, American Mathematical Society, Providence, RI, pp. 201–224.
- [28] VAN DER HOFSTAD, R. (2016). *Random Graphs and Complex Networks*, Vol. 1. Cambridge University Press.