The Annals of Probability
2014, Vol. 42, No. 1, 1–40
DOI: 10.1214/12-AOP755
© Institute of Mathematical Statistics, 2014

ASYMPTOTIC BEHAVIOR AND DISTRIBUTIONAL LIMITS OF PREFERENTIAL ATTACHMENT GRAPHS

BY NOAM BERGER, CHRISTIAN BORGS, JENNIFER T. CHAYES AND AMIN SABERI

Hebrew University, Microsoft Research, Microsoft Research and Stanford University

We give an explicit construction of the weak local limit of a class of preferential attachment graphs. This limit contains all local information and allows several computations that are otherwise hard, for example, joint degree distributions and, more generally, the limiting distribution of subgraphs in balls of any given radius k around a random vertex in the preferential attachment graph. We also establish the finite-volume corrections which give the approach to the limit.

1. Introduction. About a decade ago, it was realized that the Internet has a power-law degree distribution [2, 17]. This observation led to the so-called preferential attachment model of Barabási and Albert [4], which was later used to explain the observed power-law degree sequence of a host of real-world networks, including social and biological networks, in addition to technological ones. The first rigorous analysis of a preferential attachment model, in particular proving that it has small diameter, was given by Bollobás and Riordan [7]. Since these works there has been a tremendous amount of study, both nonrigorous and rigorous, of the random graph models that explain the power-law degree distribution; see [1] and [9] and references therein for some of the nonrigorous and rigorous work, respectively.

Also motivated by the growing graphs appearing in real-world networks, for the past five years or so, there has been much study in the mathematics community of notions of graph limits. In this context, most of the work has focused on dense graphs. In particular, there has been a series of papers on a notion of graph limits defined via graph homomorphisms [11–13, 20]; these have been shown to be equivalent to limits defined in many other senses [12, 13]. Although most of the results in this work concern dense graphs, the paper [11] also introduces a notion of graph limits for sparse graphs with bounded degrees in terms of graph homomorphisms; using expansion methods from mathematical physics, Borgs et al. [10] establishes some general results on this type of limit for sparse graphs. Another recent work [8] concerns limits for graphs which are neither dense nor sparse in the above senses; they have average degrees which tend to infinity.

Received September 2010; revised March 2012.
MSC2010 subject classifications. 60C05, 60K99.
Key words and phrases. Preferential attachment graphs, graph limits, weak local limit.

Earlier, a notion of a weak local limit of a sequence of graphs with bounded degrees was given by Benjamini and Schramm [5] (this notion was in fact already implicit in [3]). Interestingly, it is not hard to show that the Benjamini–Schramm limit coincides with the limit defined via graph homomorphisms in the case of sparse graphs of bounded degree; see [16] for yet another equivalent notion of convergent sequences of graphs with bounded degrees.

As observed by Lyons [21], the notion of graph convergence introduced by Benjamini and Schramm is meaningful even when the degrees are unbounded, provided the *average degree* stays bounded. Since the average degree of the Barabási–Albert graph is bounded by construction, it is therefore natural to ask whether this graph sequence converges in the sense of Benjamini and Schramm.

In this paper, we establish the existence of the Benjamini–Schramm limit for the Barabási–Albert graph by giving an explicit construction of the limit process, and use it to derive various properties of the limit. Our results cover the case of both uniform and preferential attachments graphs.[1] Moreover, our methods establish the finite-volume corrections which give the approach to the limit.

Our proof uses a representation, which we first introduced in [6], to analyze processes that model the spread of viral infections on preferential attachment graphs. Our representation expresses the preferential attachment model process as a combination of several Pólya urn processes. The classic Pólya urn model was of course proposed and analyzed in the beautiful work of Pólya and Eggenberger in the early twentieth century [15]; see [14] for a basic reference. Despite the fact that our Pólya urn representation is a priori only valid for a limited class of preferential attachment graphs, we give an approximating coupling which proves that the limit constructed here is the limit of a much wider class of preferential attachment graphs.

Our alternative representation contains much more independence than previous representations of preferential attachment and is therefore simpler to analyze. In order to demonstrate this, we also give a few applications of the limit. In particular, we use the limit to calculate the degree distribution and the joint degree distribution of a typical vertex with the vertex it attached to in the preferential attachment process (more precisely, a vertex chosen uniformly from the ones it attached to).

2. Definition of the model and statements of results.

2.1. *Definition of the model.* The preferential attachment graph we define generalizes the model introduced by Barabási and Albert [4] and rigorously con-

[1] Note, however, that we do not cover models exhibiting densification in the sense of Leskovec, Kleinberg and Faloutsos [19]; see [18] for a mathematical model exhibiting this phenomenon. Indeed, these models are outside the scope of convergence considered in this paper, since they have bounded diameter and growing average degree, and hence do not converge in the sense of Benjamini–Schramm.

structed in [7]. Fix an integer $m \geq 2$ and a real number $0 \leq \alpha < 1$. We will construct a sequence of graphs (G_n) (where G_n has n vertices labeled $1, \ldots, n$) as follows:

G_1 contains one vertex and no edges, and G_2 contains two vertices and m edges connecting them. Given G_{n-1} we create G_n the following way: we add the vertex n to the graph, and choose m vertices $w_1, \ldots, w_m$, possibly with repetitions, from G_{n-1}. Then we draw edges between n and each of $w_1, \ldots, w_m$. Repetitions in the sequence $w_1, \ldots, w_m$ result in multiple edges in the graph G_n.

We suggest three different ways of choosing the vertices $w_1, \ldots, w_m$. The first two ways, the independent and the conditional, are natural ways which we consider of interest, and are the two most common interpretations of the preferential attachment model. The third way, that is, the sequential model, is less natural, but is much easier to analyze because it is exchangeable, and therefore by de-Finetti's theorem (see [14]) has an alternative representation, which contains much more independence. We call this representation the Pólya urn representation because the exchangeable system we use is the Pólya urn scheme.

(1) The independent model: $w_1, \ldots, w_m$ are chosen independently of each other conditioned on the past, where for each $i = 1, \ldots, m$, we choose w_i as follows: with probability α, we choose w_i uniformly from the vertices of G_{n-1}, and with probability $1 - \alpha$, we choose w_i according to the preferential attachment rule, that is, for all $k = 1, \ldots, n-1$,

$$\mathbf{P}(w_i = k) = \frac{\deg_{n-1}(k)}{Z},$$

where Z is the normalizing constant $Z = \sum_{k=1}^{n-1} \deg_{n-1}(k) = 2m(n-2)$.

(2) The conditional model: here we start with some predetermined graph structure for the first m vertices. Then at each step, $w_1, \ldots, w_m$ are chosen as in the independent case, *conditioned* on them being different from one another.

(3) The sequential model: $w_1, \ldots, w_m$ are chosen inductively as follows: with probability α, w_1 is chosen uniformly, and with probability $1 - \alpha$, w_1 is chosen according to the preferential attachment rule, that is, for every $k = 1, \ldots, n-1$, we take $w_1 = k$ with probability $(\deg_{n-1}(k))/Z$ where as before $Z = 2m(n-2)$. Then we proceed inductively, applying the same rule, but with two modifications:

(a) When determining w_i, instead of the degree $\deg_{n-1}(k)$, we use

$$\deg'_{n-1}(k) = \deg_{n-1}(k) + \#\{1 \leq j \leq i-1 \mid w_j = k\}$$

and normalization constant

$$Z' = \sum_{k=1}^{n-1} \bigl(\deg'_{n-1}(k)\bigr) = 2m(n-2) + i - 1.$$

(b) The probability of uniform connection will be

$$\tilde{\alpha} = \alpha \frac{2m(n-1)}{2m(n-2) + 2m\alpha + (1-\alpha)(i-1)} = \alpha + O(n^{-1}) \tag{1}$$

rather than α.

We will refer to all three models as versions of the preferential attachment graph, or PA-graph, for short. Even though we consider the graph G_n as undirected, it will often be useful to think of the vertices $w_1, \dots, w_m$ as vertices which "received an edge" from the vertex n, and of n as a vertex which "sent out m edges" to the vertices $w_1, \dots, w_m$. Note in particular, that the degree of a general vertex v in G_n can be written as $m + q$, where m is the number of edges sent out by v and q is the (random) number of edges received by v.

2.2. *Pólya urn representation of the sequential model.* Our first theorem gives the Pólya urn representation of the sequential model. To state it, we use the standard notation $X \sim \beta(a, b)$ for a random variable $X \in [0, 1]$ whose density is equal to $\frac{1}{Z} x^{a-1}(1-x)^{b-1}$, where $Z = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$. We set

$$u = \frac{\alpha}{1-\alpha}.$$

Note that $u \in [0, \infty)$.

THEOREM 2.1. *Fix m, α and n. Let $\psi_1 = 1$, let $\psi_2, \dots, \psi_n$ be independent random variables with*

$$\psi_j \sim \beta\bigl(m + 2mu, (2j-3)m + 2mu(j-1)\bigr) \tag{2}$$

and let

$$\varphi_j = \psi_j \prod_{i=j+1}^{n} (1 - \psi_i), \qquad S_k = \sum_{j=1}^{k} \varphi_j \quad \text{and} \quad I_k = [S_{k-1}, S_k). \tag{3}$$

Conditioned on $\psi_1, \dots, \psi_n$, choose $\{U_{k,i}\}_{k=1,\dots,n, i=1,\dots,m}$ as a sequence of independent random variables, with $U_{k,i}$ chosen uniformly at random from $[0, S_{k-1}]$. Join two vertices j and k if $j < k$ and $U_{k,i} \in I_j$ for some $i \in \{1, \dots, m\}$ (with multiple edges between j and k if there are several such i). Denote the resulting random multi-graph by G_n.

Then G_n has the same distribution as the sequential PA-graph.

Figure 1 illustrates this theorem.

It should be noted that the $\alpha = 0$ case of the sequential model defined here differs slightly from the model of Bollobás and Riordan [7] in that they allow (self-)loops, while we do not. In fact, a minor alteration of our Pólya urn representation models their graph, and we suspect that a minor alteration of their pairing representation can model our graph.

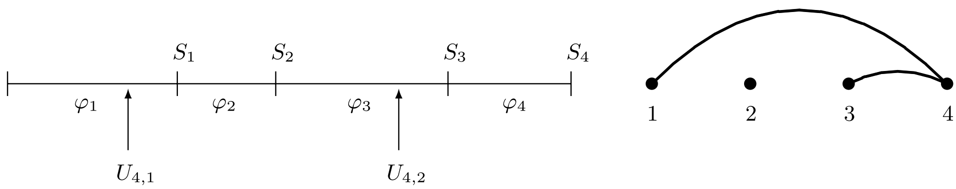

FIG. 1. *The Pólya-representation of the sequential model for $m = 2$, $n = 4$ and $k = 4$. The variables $U_{4,1}$ and $U_{4,2}$ are chosen uniformly at random from* $[0, S_3]$.

2.3. *Definition of the Pólya-point graph model.*

2.3.1. *Motivation.* The Benjamini–Schramm notion [5] of weak convergence involves the view of the graph G_n from the point of view of a "root" k_0 chosen uniformly at random from all vertices in G_n. More precisely, it involves the limit of the sequence of balls of radius $1, 2, \dots,$ about this root; see Definition 2.1 in Section 2.4 below for the details.

It turns out that for the sequential model, this limit is naturally described in terms of the random variables S_{k-1} introduced in Theorem 2.1. To explain this, it is instructive to first consider the ball of radius 1 around the random root k_0. This ball contains the m neighbors of k_0 that were born before k_0 and received an edge from k_0 under the preferential attachment rule described above, as well as a random number q_0 of neighbors that were born after k_0 and send an edge to k_0 at the time they were born. We denote these neighbors by $k_{01}, \dots, k_{0m}$ and $k_{0,m+1}, \dots, k_{0,m+q_0}$, respectively.

Let

$$\chi = \frac{1+2u}{2+2u} \quad \text{and} \quad \psi = \frac{1-\chi}{\chi} = \frac{1}{1+2u} \tag{4}$$

and note that $\frac{1}{2} \le \chi < 1$ and $0 < \psi \le 1$. As we will see, the random variables S_{k-1} behave asymptotically like $(k/n)^{\chi}$, implying in particular that the distribution of S_{k_0-1} tends to that of a random variable $x_0 = y_0^{\chi}$, where y_0 is chosen uniformly at random in $[0, 1]$. The limiting distribution of $S_{k_{01}-1}, \dots, S_{k_{0m}-1}$ turns out to be quite simple as well: in the limit these random variables are i.i.d. random variables $x_{0,i}$ chosen uniformly from $[0, x_0]$, a distribution which is more or less directly inherited from the uniform random variables $U_{k,i} \in [0, S_{k_0-1}]$ from Theorem 2.1. The limiting distribution of the random variables $S_{k_{0,m+1}-1}, \dots, S_{k_{0,m+q_0}-1}$ is slightly more complicated to derive and is given by a Poisson process in $[x_0, 1]$ with intensity

$$\gamma_0 \frac{\psi x^{\psi-1}}{x_0^{\psi}}\, dx.$$

Here γ_0 is a random "strength" which arises as a limit of the β-distributed random variable ψ_{k_0}, and is distributed according to $\Gamma(m + 2mu, 1)$. Here, as usual,

$\Gamma(a,b)$ is used to denote a distribution on $[0,\infty)$ which has density $\frac{1}{Z}x^{a-1}e^{-bx}$, with $Z=\int_0^\infty x^{a-1}e^{-bx}\,dx$.

Next, consider the branching that results from exploring the neighborhood of a random vertex in G_n in a ball of radius bigger than one. In each step of this exploration, we will find two kinds of children of the current vertex k: those which were born before k, and were attached to k at the birth of k, and those which were born after k, and were connected to k at their own birth. There are always either m or $m-1$ children of the first kind (if k was born after its parent, there will be $m-1$ such children, since one of the m edges sent out by k was sent out to k's parent; otherwise there will be m children of the first type). The number of children of the second kind is a random variable.

In the limit $n\to\infty$, this branching process leads to a random tree whose vertices, $\bar{a}$, carry three labels: a "strength" $\gamma_{\bar{a}}\in(0,\infty)$ inherited from the β-random variables ψ_k, a "position" $x_{\bar{a}}\in[0,1]$ inherited from the random variables S_{k-1} and a type which can be either L (for "left") or R (for "right"), reflecting whether the vertex k was born before or after its parent. While the strengths of vertices of type R turn out to be again $\Gamma(m+2mu,1)$-distributed, this is not the case for vertices of type L, since a vertex with higher values of ψ_k has a larger probability of receiving an edge from its child. In the limit, this will be reflected by the fact that the strength of vertices of type L is $\Gamma(m+2mu+1,1)$-distributed.

2.3.2. *Formal definition.* The main goal of the previous subsection was to give an intuition of the structure of the neighborhood of a random vertex. We will show that asymptotically, the branching process obtained by exploring the neighborhood of a random vertex k_0 in G_n is given by a random tree with a certain distribution. In order to state our main theorem, we give a formal definition of this tree.

Let F be the Gamma distribution $\Gamma(m+2mu,1)$, and let F' be the Gamma distribution $\Gamma(m+2mu+1,1)$. We define a random, rooted tree $(T,0)$ with vertices labeled by finite sequences

$$\bar{a}=(0,a_1,a_2,\dots,a_l)$$

inductively as follows:

- The root (0) has a position $x_0=y_0^\chi$, where y_0 is chosen uniformly at random in $[0,1]$. In the rest of the paper, for notational convenience, we will write 0 instead of (0) for the root.
- In the induction step, we assume that $\bar{a}=(0,a_1,a_2,\dots,a_l)$ and the corresponding variable $x_{\bar{a}}\in[0,1]$ have been chosen in a previous step. Define $(\bar{a},j)$ as $(0,a_1,a_2,\dots,a_l,j)$, $j=1,2,\dots,$ and set

$$m_-(\bar{a})=\begin{cases} m, & \text{if } \bar{a} \text{ is the root or of type } L,\\ m-1, & \text{if } \bar{a} \text{ is of type } R.\end{cases}$$

We then take

$$\gamma_{\bar{a}} \sim \begin{cases} F, & \text{if } \bar{a} \text{ is the root or of type } R, \\ F', & \text{if } \bar{a} \text{ is of type } L, \end{cases}$$

independently of everything previously sampled, choose $x_{(\bar{a},1)}, \dots, x_{(\bar{a},m_-(\bar{a}))}$ i.i.d. uniformly at random in $[0, x_{\bar{a}}]$, and $x_{(\bar{a},m_-(\bar{a})+1)}, \dots, x_{(\bar{a},m_-(\bar{a})+q_{\bar{a}})}$ as the points of a Poisson process with intensity

$$\rho_{\bar{a}}(x)\,dx = \gamma_{\bar{a}} \frac{\psi x^{\psi-1}}{x_{\bar{a}}^{\psi}}\,dx \tag{5}$$

on $[x_{\bar{a}}, 1]$ (recall that $0 < \psi \le 1$). The children of $\bar{a}$ are the vertices $(\bar{a}, 1), \dots, (\bar{a}, m_-(\bar{a})+q_{\bar{a}})$, with $(\bar{a}, 1), \dots, (\bar{a}, m_-(\bar{a}))$ called of type L, and the remaining ones called of type R.

We continue this process ad infinitum to obtain an infinite, rooted tree $(T, 0)$. We call this tree the Pólya-point graph, and the point process $\{x_{\bar{a}}\}$ the Pólya-point process.

2.4. *Main result.* We are now ready to formulate our main result, which states that in all three versions, the graph G_n converges to the Pólya-point graph in the sense of [5].

Let $\mathcal{G}$ be the set of rooted graphs, that is, the set of all pairs (G, x) consisting of a connected graph G and a designated vertex x in G, called the root. Two rooted graphs $(G, x), (G', x') \in \mathcal{G}$ are called isomorphic if there is an isomorphism from G to G' which maps x to x'. Given a finite integer r, we denote the rooted ball of radius r around x in $(G, x) \in \mathcal{G}$ by $B_r(G, x)$. We then equip $\mathcal{G}$ with the σ-algebra generated by the events that $B_r(G, x)$ is isomorphic to a finite, rooted graph (H, y) (with r running over all finite, positive integers, and (H, y) running over all finite, rooted graphs), and call (G, x) a random, rooted graph if it is a sample from a probability distribution on $\mathcal{G}$. We write $(G, x) \sim (G', x')$ if (G, x) and (G', x') are isomorphic.

DEFINITION 2.1. Given a sequence of random, finite graphs G_n, let $k_0^{(n)}$ be a uniformly random vertex from G_n. Following [5], we say that an infinite random, rooted graph (G, x) is the weak local limit of G_n if for all finite rooted graphs (H, y) and all finite r, the probability that $B_r(G_n, k_0^{(n)})$ is isomorphic to (H, y) converges to the probability that $B_r(G, x)$ is isomorphic to (H, y).

The main result of the paper is the following theorem.

THEOREM 2.2. *The weak local limit of the all three variations of the preferential attachment model is the Pólya-point graph.*

Recently, and independently of our work, Rudas et al. [22], studied the random tree resulting from the preferential attachment model when $m = 1$. They derived the asymptotic distribution of the subtree under a randomly selected vertex which implies the Benjamini–Schramm limit. Note that when $m = 1$, there is no distinction between the independent, conditional and sequential models.

As alluded to before, the points $x_{\bar{a}}$ of the Pólya-point process represent the random variables S_{k-1} of the vertices in G_n, which in turn behave like $(k/n)^{\chi}$ as $n \to \infty$. The variable $y_{\bar{a}} = x_{\bar{a}}^{1/\chi}$ thus represents the birth-time of the corresponding vertex in G_n. This is made precise in the following corollary to the proof of Theorem 2.2. As the theorem, the corollary holds for all three versions of the Preferential Attachment model.

COROLLARY 2.3. *Given $r < \infty$ and $\varepsilon > 0$ there exists a $n_0 < \infty$ such that for $n \geq n_0$, there exists a coupling μ between a sample T of the Pólya-point and an ensemble $\{G_n, v_0\}$ where G_n has the distribution of the preferential attachment graph of size n, and v_0 is a uniformly chosen vertex of G_n, satisfying: with μ probability at least $1 - \varepsilon$, there exists an isomorphism $\bar{a} \mapsto k_{\bar{a}}$ from the ball of radius r about 0 in $(T, 0)$ into the ball of radius r about v_0 in G_n, with the property that*

$$\left| y_{\bar{a}} - \frac{k_{\bar{a}}}{n} \right| \leq \varepsilon$$

for all $\bar{a}$ with distance at most r from the root in $(T, 0)$. Here $y_{\bar{a}}$ is defined as $y_{\bar{a}} = x_{\bar{a}}^{1/\chi}$.

The numerator $x_{\bar{a}}^{\psi} = y_{\bar{a}}^{1-\chi}$ in (5) thus expresses the fact that in the preferential attachment process, earlier vertices are more likely to attract many neighbors than later vertices.

2.5. *Subgraph frequencies.* A natural question concerning a sequence of growing graphs (G_n) is the question of how often a small graph F is contained in G_n as a subgraph. This question can be formalized in several ways, for example, by considering the number of homomorphisms from F into G_n, or the number of injective homomorphism, or the number of times F is contained in G_n as an *induced* subgraph.

For graph sequences with bounded degrees, this leads to an alternative notion of convergence, by defining sequence of graphs to be convergent if the homomorphism density $t(F, G_n)$—defined as the number of homomorphisms from F into G_n divided by the number of vertices in G_n—converges for all finite connected graphs F [10, 11]. Indeed, for sequences of graphs G_n whose degree is bounded uniformly in n, this notion can easily be shown to be equivalent to the notion introduced by Benjamini and Schramm; moreover, the corresponding notions involving the number of injective homomorphisms, or the number of induced subgraphs, are equivalent as well; see [11], Section 2.2 for formulas expressing these various numbers in terms of each other.

But for graphs with growing maximal degree, this equivalence does not hold in general. Indeed, consider a sequence of graphs with uniformly bounded degrees, augmented by a vertex of degree $n^{1/2}$. Such a vertex does not change the notion of convergence introduced by Benjamini and Schramm; however, the number of homomorphisms from a star with 3 legs into this graph sequence grows like $n^{3/2}$, implying that the homomorphism density diverges.

To overcome this difficulty, we will consider maps Φ from $V(F)$, the vertex set of F, into $V(G_n)$, the vertex set of G_n which in addition to being homormorphisms also preserve degrees. More explicitly, given a graph F and a map $\mathbf{n}: V(F) \to \{0, 1, 2, \ldots\}$, we define $\mathrm{inj}(F, \mathbf{n}; G_n)$ as the number of injective maps $\Phi: V(F) \to V(G_n)$ such that:

(1) If $ij \in E(F)$, then $\Phi(i)\Phi(j) \in E(G_n)$;
(2) $d_{\Phi(i)}(G_n) = d_i(F) + n(i)$ for all $i \in V(F)$,

where $E(F)$ denotes the set of edges in F, and $d_i(F)$ denotes the degree of the vertex i in F.

The following lemma is due to Laci Lovasz.

LEMMA 2.4. *Let $D < \infty$, and let G_n be a sequence of graphs that converges in the sense of Benjamini and Schramm. Then the limit*

$$\hat{t}(F, \mathbf{n}) = \lim_{n\to\infty} \frac{1}{|V(G_n)|} \mathrm{inj}(F, \mathbf{n}; G_n)$$

exists for all finite connected graphs F and all maps $\mathbf{n}: V(F) \to \{0, 1, 2, \ldots\}$.

As stated, the lemma refers to sequences of deterministic graphs. For sequences of random graphs, its proof gives convergence of the expected number of the subgraph frequencies $\frac{1}{|V(G_n)|} \mathrm{inj}(F, \mathbf{n}; G_n)$. To prove convergence in probability for these frequencies, a little more work is needed. For the case of preferential attachment graphs, we do this in Section 5.4.3, together with an explicit calculation of the actual values of these numbers.

REMARK 2.5. When G_n has multiple edges, the definition of $\mathrm{inj}(F, \mathbf{n}; G_n)$ has to be modified. There are a priory several possible definitions; motivated by the notions introduced in [11] we chose the definition

$$\mathrm{inj}(F, \mathbf{n}; G_n) = \sum_{\Phi} \prod_{ij\in E(F)} m_{\Phi(i)\Phi(j)}(G_n)^{m_{ij}(H)},$$

where the sum goes over injective maps $\Phi: V(H) \to V(G_n)$ obeying condition (2) above with $d_i(H)$ and $d_{\Phi(i)}(G_n)$ denoting degrees counting multiplicities, and where $m_{ij}(H)$ is the multiplicity of the edge ij in H [and similarly for $m_{\Phi(i)\Phi(j)}(G_n)$]. With this definition, the above lemma holds for graphs with multiple edges as well.

3. Proof of weak distributional convergence for the sequential model. In this section we prove that the sequential model converges to the Pólya-point tree.

3.1. *Pólya urn representation of the sequential model.* In the early twentieth century, Pólya proposed and analyzed the following model known as the Pólya urn model; see [14]. The model is described as follows. We have a number of urns, each holding a number of balls, and at each step, a new ball is added to one of the urns. The probability that the ball is added to urn i is proportional to $N_i + u$ where N_i is the number of balls in the ith urn and u is a predetermined parameter of the model.

Pólya showed that this model is equivalent to another process as follows. For every i, choose at random a parameter (which we call "strength" or "attractiveness") p_i, and at each step, *independently* of our decision in previous steps, put the new ball in urn i with probability p_i. Pólya specified the distribution (as a function of u and the initial number of balls in each urn) for which this mimics the urn model. A particularly nice example is the case of two urns, each starting with one ball and $u = 0$. Then p_1 is a uniform [0, 1] variable, and $p_2 = 1 - p_1$. Pólya showed that for general values of u and $\{N_i(0)\}$, the values of $\{p_i\}$ are determined by the β-distribution with appropriate parameters.

It is not hard to see that there is a close connection between the preferential attachment model of Barabási and Albert and the Pólya urn model in the following sense: every new connection that a vertex gains can be represented by a new ball added in the urn corresponding to that vertex.

To derive this representation, let us consider first a two-urn model, with the number of balls in one urn representing the degree of a particular vertex k, and the number of balls in the other representing the sum of the degrees of the vertices $1, \dots, k-1$. We will start this process at the point when $n = k$ and k has connected to precisely m vertices in $\{1, \dots, k-1\}$. Note that at this point, the urn representing the degree of k has m balls, while the other one has $(2k-3)m$ balls.

Consider a time in the evolution of the preferential attachment model when we have $n - 1 \geq k$ old vertices, and $i - 1$ edges between the new vertex n and $\{1, \dots, k-1\}$. Assume that at this point the degree of k is d_k, and the sum of the degrees of $1, \dots, k-1$ is $d_{<k}$. At this point, the probability that the ith edge from n to $\{1, \dots, n-1\}$ is attached to k is

$$\begin{aligned}\tilde{\alpha}\frac{1}{n-1} + (1-\tilde{\alpha})\frac{d_k}{2m(n-2)+(i-1)}\\ = \frac{2m\alpha + (1-\alpha)d_k}{2m(n-2)+2m\alpha+(1-\alpha)(i-1)},\end{aligned} \tag{6}$$

while the probability that it is attached to one of the nodes $1, \dots, k-1$ is

$$\begin{aligned}(7)\qquad &\tilde{\alpha}\frac{k-1}{n-1} + (1-\tilde{\alpha})\frac{d_{<k}}{2m(n-2)+(i-1)} \\ &\quad = \frac{2m\alpha + (1-\alpha)d_{<k}}{2m(n-2)+2m\alpha+(1-\alpha)(i-1)}.\end{aligned}$$

Thus, conditioned on connecting to $\{1, \dots, k\}$, the probability that the ith edge from n to $\{1, \dots, n-1\}$ is attached to k is

$$\frac{1}{Z}(2mu + d_k),$$

while the conditional probability that it is attached to one of the nodes $1, \dots, k-1$ is

$$\frac{1}{Z}(2mu(k-1) + d_{<k}),$$

where Z is an appropriate normalization constant. Note that the constant $\tilde{\alpha}$ in (1) was chosen in such a way that the factor u appearing in these expressions does not depend on i, which is crucial to guaranty exchangeability.

Taking into account that the two urns start with m and $(2k-3)m$ balls, respectively, we see that the evolution of the two bins is a Pólya urn with strengths ψ_k and $1-\psi_k$, where $\psi_k \sim B_k = \beta(m + 2mu, (2k-3)m + 2mu(k-1))$.

PROOF OF THEOREM 2.1. Using the two urn process as an inductive input, we can now easily construct the Pólya graph defined in Theorem 2.1. Indeed, let $X_t \in \{1, 2, \dots, \lceil \frac{t}{m} \rceil\}$ be the vertex receiving the tth edge in the sequential model (the other endpoint of this edge being the vertex $\lceil \frac{t}{m} \rceil + 1$). For $t \le m$, X_t is deterministic (and equal to 1), but starting at $t = m+1$, we have a two-urn model, starting with m balls in each urn. As shown above, the two urns can be described as Pólya-urns with strengths $1-\psi_2$ and ψ_2. Once $t > 2m$, X_t can take three values, but conditioned on $X_t \le 2$, the process continues to be a two-urn model with strengths $1-\psi_2$ and ψ_2. To determine the probability of the event that $X_t \le 2$, we now use the above two-urn model with $k = 3$, which gives that the probability of the event $X_t \le 2$ is $1-\psi_3$, at least as long as $t \le 3m$. Combining these two-urn models, we get a three-urn model with strengths $(1-\psi_2)(1-\psi_3)$, $\psi_2(1-\psi_3)$ and ψ_3. Again, this model remains valid for $t > 3m$, as long as we condition on $X_t \le 3$.

Continuing inductively, we see that the sequence X_t evolves in stages:

- For $t = 1, \dots, m$, the variable X_t is deterministic: $X_t = 1$.
- For $t = m+1, \dots, 2m$, the distribution of $X_t \in \{1, 2\}$ is described by a two-urn model with strengths $1-\psi_2$ and ψ_2, where $\psi_2 \sim B_2$.

- In general, for $t = m(k-1)+1, \dots, km$, the distribution of $X_t \in \{1, \dots, k\}$ is described by a k-urn model with strengths

$$\varphi_j^{(k)} = \psi_j \prod_{i=j+1}^{k} (1-\psi_i), \qquad j = 1, \dots, k. \tag{8}$$

  Here $\psi_k \sim B_k$ is chosen at the beginning of the kth stage, independently of the previously chosen strengths $\psi_1, \dots, \psi_{k-1}$ (for convenience, we set $\psi_1 = 1$).

Note that the random variables $\varphi_j^{(k)}$ can be expressed in terms of the random variables introduced in Theorem 2.1 as follows: by induction on k, it is easy to show that

$$S_k = \prod_{j=k+1}^{n} (1-\psi_k). \tag{9}$$

This implies that

$$\varphi_j^{(k)} = \frac{\psi_j}{S_k},$$

which relates the strengths $\varphi_j^{(k)}$ to the random variables defined in Theorem 2.1, and shows that the process derived above is indeed the process given in the theorem. $\square$

In order to apply Theorem 2.1, we will use two technical lemmas, whose proofs will be deferred to a later section. The first lemma states a law of large numbers for the random variables S_k.

LEMMA 3.1. *For every ε there exist $K < \infty$ such that for $n \geq K$, we have that with probability at least $1-\varepsilon$,*

$$\max_{k \in \{1,\dots,n\}} \left| S_k - \left(\frac{k}{n}\right)^{\chi} \right| \leq \varepsilon$$

and

$$\max_{k \in \{K,\dots,n\}} \left| S_k - \left(\frac{k}{n}\right)^{\chi} \right| \leq \varepsilon \left(\frac{k}{n}\right)^{\chi}.$$

The second lemma concerns a coupling of the sequence $\{\psi_k\}_{k\geq 1}$ and an i.i.d. sequence of Γ-random variables $\{\chi_k\}_{k\geq 1}$, where $\chi_k \sim \Gamma(m+2mu, 1)$. To describe the coupling, we define a sequence of functions $f_k : [0,\infty) \to [0,1)$ by

$$\mathbf{P}\big(\psi_k \leq f_k(x)\big) = \mathbf{P}(\chi_k \leq x). \tag{10}$$

Then $f_k(\chi_k)$ has the same distribution as ψ_k, implying that $(\{\chi_k\}_{k\geq 1}, \{f_k(\chi_k)\}_{k\geq 1})$ defines a coupling between $\{\chi_k\}_{k\geq 1}$ and $\{\psi_k\}_{k\geq 1}$.

LEMMA 3.2. *Let f_k be as in (10), and let $\{\chi_k\}_{k\geq 1}$ i.i.d. random variables with distribution $\Gamma(m+2mu,1)$. Given $\varepsilon>0$ there exist a $K<\infty$ so that the following holds*:

(i) *With probability at least $1-\varepsilon$*,

$$\chi_k \leq \log^2 k \qquad \text{for all } k \geq K; \tag{11}$$

(ii) *For $k\geq K$ and $x\leq \log^2 k$*,

$$\frac{1-\varepsilon}{2mk(1+u)}x \leq f_k(x) \leq \frac{1+\varepsilon}{2mk(1+u)}x. \tag{12}$$

We defer the proof of Lemmas 3.1 and 3.2 to Section 3.6.

3.2. *The exploration tree of G_n.* Let $K_r=K_r(G_n,k_0)$ be the set of vertices in G_n which have distance at most r from the random root k_0, and let $\hat{B}_r(G_n,k_0)$ be the graph on K_r that contains all edges in G_n for which at least one endpoint has distance $\leq r$ from k_0. When proving that the preferential model converges to the Pólya-point graph, we will use the notion of convergence given in Definition 2.1, but instead of the standard ball of radius r, we will use the modified ball $\hat{B}_r(G_n,k_0)$. (It is obvious that this definition is equivalent.)

We will prove our results by induction on r, using the exploration procedure outlined in Section 2.3.1 in the inductive step. To this end, it will be convenient to endow the rooted graph (G_n,k_0) with a structure which is similar to the one defined for the Pólya-point graph. More precisely, we will inductively define a rooted tree $(T_r^{(n)},0)$ on sequence of integers $\bar{a}=(0,a_1,a_2,\dots,a_l)$, and a homomorphism

$$\mathbf{k}^{(r)}: T_r^{(n)} \to \hat{B}_r(G_n,k_0): \bar{a}\mapsto k_{\bar{a}}$$

as follows.

We start our inductive definition by mapping 0 into a vertex k_0 chosen uniformly at random from the vertex set $\{1,\dots,n\}$ of G_n. Given a vertex $\bar{a}=(0,a_1,a_2,\dots,a_l)\in T_r^{(n)}$ and its image $k_{\bar{a}}$ in G_n, let $d_{\bar{a}}$ be the degree of $k_{\bar{a}}$ in G_n, and let $k_{\bar{a}_-},k_1,\dots,k_{d_{\bar{a}}-1}$ be the neighbors of $k_{\bar{a}}$ in G_n, where $\bar{a}_-=(0,a_1,a_2,\dots,a_{l-1})$. Recalling that edges were created one by one during the sequential preferential attachment process, we order $k_1,\dots,k_{d_{\bar{a}}-1}$ in such a way that for all $i=1,\dots,d_{\bar{a}}-2$, the edge $(k_{\bar{a}},k_i)$ was born before the edge $(k_{\bar{a}},k_{i+1})$. We then define the children of $\bar{a}$ to be the points $(\bar{a},1),\dots,(\bar{a},d_{\bar{a}}-1)$. This defines $T_{r+1}^{(n)}$. The map $\mathbf{k}^{(r+1)}$ is the extension of $\mathbf{k}^{(r)}$ which maps $(\bar{a},1),\dots,(\bar{a},d_{\bar{a}}-1)$ to the vertices $k_1,\dots,k_{d_{\bar{a}}-1}$, respectively. We call a vertex $(\bar{a},i)$ early or of type L if $k_i<k_{\bar{a}_-}$ and late or of type R otherwise. Note that the root and vertices of type L have m children of type L, while vertices of type R have $m-1$ children of type L.

To make the dependence on G_n explicit, we often use the notation $T_r(G_n)$ for the tree $T_r^{(n)}$, and the notation $\mathbf{k}^{(r)}(G_n)$ for the map $\mathbf{k}^{(r)}$. Note that $\mathbf{k}^{(r)}$ does not, in general, give a graph isomorphism between $T_r^{(n)}$ and $\hat{B}_r(G_n, k_0)$. But if the map is injective when restricted to $T_r^{(n)}$, it is a graph isomorphism. To prove Theorem 2.2, it is therefore enough to show that for all r, the map $\mathbf{k}^{(r)}$ is injective and the tree $T_r^{(n)}$ converges in distribution to T_r, the ball of radius r in the Pólya-point graph $(T, 0)$.

3.3. *Regularity properties of the Pólya-point process.* In order to prove Theorem 2.2, we will use some simple regularity properties of the Pólya-point process.

Recall the definition of the Pólya-point graph $(T, 0)$ and the Pólya-point process $\{x_{\bar{a}}\}$ from Section 2.3.2, as well as the notation $\rho_{\bar{a}}(x)\,dx$ for the intensity defined in (5). As usual, we define the height of a vertex $\bar{a} = (0, a_1, a_2, \ldots, a_l)$ in T as its distance l from the root. We denote the rooted subtree of height r in $(T, 0)$ by $(T_r, 0)$.

LEMMA 3.3. *Fix $0 \le r < \infty$ and $\varepsilon > 0$. Then there are constants $\delta > 0$, $C < \infty$, $K < \infty$ and $N < \infty$ such that with probability at least $1 - \varepsilon$, we have that*:

- *$x_{\bar{a}} \ge \delta$ for all vertices $\bar{a}$ in T_r;*
- *$\gamma_{\bar{a}} \le C$;*
- *$\rho_{\bar{a}}(\cdot) \le K$;*
- *$|T_r| \le N$.*

PROOF. The proof of the lemma is easily obtained by induction on r. We leave it to the reader. $\square$

COROLLARY 3.4. *For all $\varepsilon > 0$ and all $r < \infty$ there is a constant $\delta > 0$ such that with probability at least $1 - \varepsilon$, we have*

$$\min_{\substack{\bar{a}, \bar{b} \in T_r \\ \bar{a} \neq \bar{b}}} |x_{\bar{b}} - x_{\bar{a}}| \ge \delta.$$

PROOF. This is an immediate consequence of the continuous nature of the random variables $x_{\bar{a}}$ and the statements of Lemma 3.3. $\square$

3.4. *The neighborhood of radius one.* Before proving our main theorem, Theorem 2.2, for the sequential model, we establish the following lemma, which will serve as the base in an inductive proof of our main theorem.

LEMMA 3.5. *Let G_n be the sequential preferential attachment graph, let k_0 be chosen uniformly at random in $\{1, \ldots, n\}$ and let $k_{0,1}, \ldots, k_{0,m+q_0}$ be the neighbors of k_0, ordered as in Section 3.2 by the birth times of the edges $\{k_0, k_{0,i}\}$. Then (G_n, k_0) and the Pólya-point process $\{x_{\bar{a}}\}$ can be coupled in such a way that for all $\varepsilon > 0$ there are constants $C, N < \infty$, $\delta > 0$ and $n_0 < \infty$ such that for $n \ge n_0$, with probability at least $1 - \varepsilon$, we have that*:

(i) $T_1 \cong T_1(G_n)$ *and* $|T_1(G_n)| \le N$;
(ii) $|x_{\bar a} - S_{k_{\bar a}-1}| \le \varepsilon$ *for all* $\bar a \in T_1$;
(iii) $k_0, k_{0,1}, \dots, k_{0,m+q_0}$ *are pairwise distinct and* $k_{\bar a} \ge \delta n$ *for all* $\bar a \in T_1$;
(iv) $\chi_{k_{\bar a}} = \gamma_{\bar a} \le C$ *for all* $\bar a \in T_1$.

PROOF. (i)–(ii): We start by proving the first two statements. Choose y_0 uniformly at random in [0, 1], let $x_0 = y_0^\chi$ and let $x_{0,1}, \dots, x_{0,m+q_0'}$ be the positions of the children of 0 in $(T, 0)$. Define $k_0 = \lceil n y_0 \rceil$, so that k_0 is distributed uniformly in $\{1, \dots, n\}$, and for $i = 1, \dots, m$, define $k_{0,i}$ by

$$S_{k_{0,i}-1} \le \frac{x_{0,i}}{x_0} S_{k_0-1} < S_{k_{0,i}}.$$

By Theorem 2.1 and the observation that $U_{k_0,1} = \frac{x_{0,1}}{x_0}, \dots, U_{k_0,m} = \frac{x_{0,m}}{x_0}$ are i.i.d. random variables chosen uniformly at random from [0, 1], we have that indeed, with large probability, $k_{0,1}, \dots, k_{0,m}$ are close enough to the $x_{0,i}$'s.

Indeed, given $\varepsilon > 0$ choose δ, C, K and N in such a way that the statements of Lemma 3.3 and Corollary 3.4 hold for $r = 1$, and let $\varepsilon' = \min\{\varepsilon, \delta/4\}$. By Lemma 3.1 there exists a constant $n_0 < \infty$ such that for $n \ge n_0$, we have that

$$|S_{k_0-1} - x_0| \le \varepsilon' \quad \text{and} \quad |S_{k_{0,i}-1} - x_{0,i}| \le \varepsilon' \qquad \text{for all } i = 1, \dots, m \tag{13}$$

with probability at least $1 - 2\varepsilon$.

To understand the limiting distribution of the remaining neighbors, $k_{0,m+1}, \dots, k_{0,m+q_0}$, of k_0, we observe that conditioned on the random variables $\psi_1, \dots, \psi_n$, each vertex $k > k_0$ has m independent chances of being connected to k_0, corresponding to the m independent events $\{X_{k,i} = k_0\}$, $i = 1, \dots, m$, where we used the shorthand $X_{k,i}$ for the interval containing the endpoint of the ith edge sent out from k (it is related to the random variables X_t introduced in the proof of Theorem 2.1 via $X_{k,i} = X_{(k-2)m+i}$). Let

$$P_{k\to k_0} = \varphi_{k_0} \frac{1}{S_{k-1}} = \frac{S_{k_0}}{S_{k-1}} \psi_{k_0} \tag{14}$$

be the probability of the event $\{X_{k,i} = k_0\}$, and let $N_{y_0}(y) = \sum_{i=1}^m \sum_{k=k_0}^{\lceil ny \rceil} \mathbb{I}(X_{k,i} = k_0)$ where $\mathbb{I}(A)$ is the indicator function of the event A. We want to show that $N_{y_0}(\cdot)$ converges to a Poisson process on $[y_0, 1]$.

By Lemma 3.3, we have that $k_0 \ge n x_0 \ge n\delta$ with probability at least $1 - \varepsilon$, which allows us to apply Lemmas 3.1 and 3.2 to show that for n large enough, with probability at least $1 - 2\varepsilon$, we have

$$\hat P_{k\to k_0}(1-\varepsilon) \le P_{k\to k_0} \le (1+\varepsilon)\hat P_{k\to k_0}$$

where

$$\hat P_{k\to k_0} = \frac{1}{nm} \frac{\chi_{k_0}}{2(1+u)} \frac{n}{k_0} \left(\frac{k_0}{k}\right)^\chi.$$

For $y > y_0$, let $\hat{N}_{y_0}(y) = \sum_{i=1}^{m} \sum_{k=k_0}^{\lceil ny \rceil} \hat{Y}_{k\to k_0}^{(i)}$ where $\{\hat{Y}_{k\to k_0}^{(i)}\}$ are independent random variables such that $\hat{Y}_{k\to k_0}^{(i)} = 1$ with probability $\hat{P}_{k\to k_0}$ and $\hat{Y}_{k\to k_0}^{(i)} = 0$ with probability $1 - \hat{P}_{k\to k_0}$. It follows from standard results on convergence to Poisson processes (and the fact that γ_0 has the same distribution as χ_{k_0}) that $\hat{N}_{y_0}(\cdot)$ converges weakly to a Poisson process with density $\frac{\gamma_0}{2(u+1)y_0}(\frac{y_0}{y})^\chi$ on $[y_0, 1]$. A change of variables from y to $x = y^\chi$ now leads to the Poisson process with density

$$\frac{\gamma_0}{2(1+u)\chi} \frac{x^{\psi-1}}{x_0^\psi} = \gamma_0 \frac{\psi x^{\psi-1}}{x_0^\psi}$$

on $[x_0, 1]$. Combined with a last application of Lemma 3.1 to bound the difference between $S_{k_{0,i}-1}$ and $(k_{0,i}/n)^\chi$, this proves that $x_{0,m+1}, \dots, x_{m+q_0'} \in [x_0, 1]$ and $k_{0,m+1}, \dots, k_{0,m+q_0}$ can be coupled in such a way that for n large enough, with probability at least $1 - 3\varepsilon$, we have that $q_0 = q_0' \le Q = N - m - 1$, $\chi_{k_0} = \gamma_0 \le C$ and

$$|x_{0,i} - S_{k_{0,i}-1}| \le \varepsilon' \qquad \text{for } i = m+1, \dots, m+q_0. \tag{15}$$

Since $\varepsilon > 0$ was arbitrary, this completes the proof of the first two statements of the lemma.

(iii) To prove the third statement, we use bounds (13) and (15), and a final application of Lemma 3.1, to establish the existence of two constants $\delta' > 0$ and $n_0' < \infty$ such that for $n \ge n_0'$, with probability at least $1 - 4\varepsilon$,

$$k_{\bar{a}} \ge \delta' n \qquad \text{for all } \bar{a} \in T_1(G_n) \tag{16}$$

and

$$|k_{\bar{a}} - k_{\bar{b}}| \ge \delta' n \qquad \text{for all } \bar{a}, \bar{b} \in T_1(G_n) \text{ with } \bar{a} \ne \bar{b},$$

implying in particular that $k_0, k_{0,1}, \dots, k_{0,m+q_0}$ are pairwise distinct.

(iv) To prove the last statement, let us assume that $\gamma_0 \le C$, and that $k_{0,1}, \dots, k_{0,m+q}$ are pairwise distinct, with $k_{0,i} < k_0$ for $i \le m$, $k_{0,i} > k_0$ for $i > m$, $\min k_{0,i} \ge n\delta'$ and $q \le Q$. Let A be the event that we have chosen k_0 as the uniformly random vertex and that the neighbors of k_0 are the vertices $k_{0,1}, \dots, k_{0,m+q}$. Let χ^{A,γ_0} be the collection of random variables $\{\chi_k\}_{k\ne k_0}$ conditioned on $\chi_{k_0} = \gamma_0$ and A. We will want to show that χ^{A,γ_0} can be coupled to a collection of independent random variables $\{\hat{\chi}_k\}_{k\ne k_0}$ such that $\chi^{A,\gamma_0} = \{\hat{\chi}_k\}_{k\ne k_0}$ with probability at least $1 - \varepsilon$, and

$$\hat{\chi}_k \sim \begin{cases} F_k', & \text{if } k \in \{k_{0,1}, \dots, k_{0,m}\}, \\ F_k, & \text{otherwise.} \end{cases} \tag{17}$$

Let $\rho(\cdot \mid A, \chi_{k_0})$ be the density of the (multi-dimesional) random variable χ^{A,γ_0}, and let $\mathbf{P}(\cdot)$ be the joint distribution of G_n and the random variables $\chi_1, \dots, \chi_n$. By Bayes's theorem,

$$\rho(\cdot \mid A, \chi_{k_0} = \gamma_0) = \frac{\mathbf{P}(A \mid \cdot, \chi_{k_0} = \gamma_0)}{\mathbf{P}(A \mid \chi_{k_0} = \gamma_0)} \rho_0(\cdot), \tag{18}$$

where ρ_0 is the original density of the random variables $\{\chi_k\}_{k\neq k_0}$ (we denote the corresponding probability distribution and expectations by P_0 and E_0, resp.).

We thus have to determine the probability of A conditioned on $\chi_1,\dots,\chi_n$. With the help of Theorem 2.1, this probability is easily calculated, and is equal to

$$\begin{aligned}\mathbf{P}(A\mid\{\chi_k\}) &= m!\prod_{i=1}^m P_{k_0\to k_{0,i}}\prod_{j=1}^q mP_{k_{0,m+j}\to k_0}(1-P_{k_{0,m+j}\to k_0})^{m-1}\\ &\qquad\times\prod_{k>k_0:k\notin\{k_{0,m+1},\dots,k_{0,m+q}\}}(1-P_{k\to k_0})^m\\ &= m!\prod_{i=1}^m P_{k_0\to k_{0,i}}\prod_{j=1}^q\frac{mP_{k_{0,m+j}\to k_0}}{1-P_{k_{0,m+j}\to k_0}}\prod_{k>k_0}(1-P_{k\to k_0})^m,\end{aligned}$$

where $P_{k\to k'}$ is the conditional probability defined in (14). By Lemma 3.1, this implies that given any $\varepsilon'>0$, we can find $n_0<\infty$ such that for $n\geq n_0$, we have that with probability at least $1-\varepsilon'$ with respect to P_0,

$$\begin{aligned}&(1-\varepsilon')\mathbf{P}(A\mid\{\chi_k\})\\ &\quad\leq m!\left(\prod_{i=1}^m\psi_{k_{0,i}}\left(\frac{k_{0,i}}{k_0}\right)^\chi\prod_{j=m+1}^{m+q}m\psi_{k_0}\left(\frac{k_0}{k_{0,j}}\right)^\chi\right)\exp\left(-m\psi_{k_0}\sum_{k>k_0}\left(\frac{k_0}{k}\right)^\chi\right)\\ &\quad\leq(1+\varepsilon')\mathbf{P}(A\mid\{\chi_k\}).\end{aligned}$$

To estimate $\mathbf{P}(A\mid\chi_{k_0})=E_0[\mathbf{P}(A\mid\{\chi_k\})]$, we combined this bound with the deterministic upper bound

$$\begin{aligned}&\mathbf{P}(A\mid\{\chi_k\})\\ &\quad\leq m!\prod_{i=1}^m P_{k_0\to k_{0,i}}\prod_{j=1}^q mP_{k_{0,m+j}\to k_0}\leq\frac{1}{n}(m\psi_{k_0})^q\prod_{i=1}^m\psi_{k_{0,i}}\\ &\quad\leq C'm!\left(\prod_{i=1}^m\psi_{k_{0,i}}\left(\frac{k_{0,i}}{k_0}\right)^\chi\prod_{j=m+1}^{m+q}m\psi_{k_0}\left(\frac{k_0}{k_{0,j}}\right)^\chi\right)\exp\left(-m\psi_{k_0}\sum_{k>k_0}\left(\frac{k_0}{k}\right)^\chi\right),\end{aligned}$$

where $C'=(\delta')^{-(m+Q)}\sup_{n\geq 1}e^{mnf_{\delta'n}(C)}$.

These bounds imply that given any $\varepsilon'>0$, we can find an $n_0<\infty$ such that for $n\geq n_0$, with probability at least $1-\varepsilon'/2$ with respect to P_0, we have

$$\sqrt{1-\varepsilon'}\prod_{i=1}^m\frac{\psi_{k_{0,i}}}{E_0(\psi_{k_{0,i}})}\leq\frac{\mathbf{P}(A\mid\{\chi_k\})}{\mathbf{P}(A\mid\chi_{k_0})}\leq\sqrt{1+\varepsilon'}\prod_{i=1}^m\frac{\psi_{k_{0,i}}}{E_0(\psi_{k_{0,i}})}.$$

With the help of Lemma 3.2, this shows that for n large enough, with probability at least $1-\varepsilon'$, we have

$$(1-\varepsilon')\prod_{i=1}^m\frac{\chi_{k_{0,i}}}{E_0(\chi_{k_{0,i}})}\leq\frac{\mathbf{P}(A\mid\{\chi_k\})}{\mathbf{P}(A\mid\chi_{k_0})}\leq(1+\varepsilon')\prod_{i=1}^m\frac{\chi_{k_{0,i}}}{E_0(\chi_{k_{0,i}})}.$$

Recalling (18) and the definition of the random variables $\{\hat{\chi}_k\}_{k\neq k_0}$, we therefore have shown that with probability at least $1-\varepsilon'$ with respect to P_0,

$$(1-\varepsilon')\hat{\rho}(\cdot) \leq \rho(\cdot \mid A, \chi_{k_0}=\gamma_0) \leq (1+\varepsilon')\hat{\rho}(\cdot), \tag{19}$$

where $\hat{\rho}$ is the density of the random variables $\{\hat{\chi}_k\}_{k\neq k_0}$. (We denote the corresponding product measure by $\hat{P}$.)

To continue, we need to transform statements which happen with high probability with respect to P_0 into statements which happen with high probability with respect to $\hat{P}$. To this end, we consider the general case of two probability measures μ and ν such that ν is absolutely continuous with respect to μ, $\nu = f\mu$ for some nonnegative function $f \in L_2(\mu)$. Let Ω_0 be an event which happens with probability $1-\varepsilon'$ with respect to μ. Then

$$\nu(\Omega_0^c) = \int f 1_{\Omega_0^c} \leq \sqrt{E_\mu(f^2)\mu(\Omega_0^c)} = \sqrt{\varepsilon' E_\mu(f^2)}, \tag{20}$$

implying that Ω_0 happens with probability at least $1-\sqrt{\varepsilon' E_\mu(f^2)}$ with respect to ν.

Applying this bound to the probability measures P_0 and $\hat{P}$, we see that bound (19) holds with probability at least $1-\sqrt{2\varepsilon'}$ with respect to $\hat{P}$, provided n (and hence $k_{0,1},\dots,k_{0,m}$) is large enough. Using this fact, one then easily shows that

$$\|\hat{\rho} - \rho(\cdot \mid A, \chi_{k_0}=\gamma_0)\|_1 \leq 2\varepsilon' + 2\sqrt{2\varepsilon'}.$$

Choosing ε' sufficiently small ($\varepsilon' = \varepsilon^2/32$ is small enough), we see that the right-hand side can be bounded by ε, which proves that χ^{A,γ_0} and $\{\hat{\chi}_k\}_{k\neq k_0}$ can be coupled in such a way that they are equal with probability at least $1-\varepsilon$, as required. $\square$

3.5. *Proof of convergence for the sequential model.* In this section we show that the sequential model converges to the Pólya-point graph. Indeed, we prove slightly more, namely the following proposition:

PROPOSITION 3.6. *Given $\varepsilon > 0$ and $r < \infty$, there are constants $C, N < \infty$, $\delta > 0$ and $n_0 < \infty$ such that for $n \geq n_0$, the rooted sequential attachment graph (G_n, k_0) and the Pólya-point process $\{x_{\bar{a}}\}$ can be coupled in such a way that with probability at least $1-\varepsilon$, the following holds*:

(1) $T_r(G_n) \cong T_r$ *and* $|T_r(G_n)| \leq N$;
(2) $|x_{\bar{a}} - S_{k_{\bar{a}}-1}| \leq \varepsilon$ *for all* $\bar{a} \in T_r$;
(3) $\mathbf{k}^{(r)}(G_n)$ *is injective, and* $k_{\bar{a}} \geq \delta n$ *for all* $\bar{a} \in T_r$;
(4) $\gamma_{\bar{a}} = \chi_{k_{\bar{a}}} \leq C$ *for all* $\bar{a} \in T_r$.

PROOF. For $r = 1$, this follows from Lemmas 3.5 and 3.3.

Assume by induction that the lemma holds for $r < \infty$, and fix T_r, $\mathbf{k}^{(r)}(G_n)$, $\{x_{\bar{a}}\}_{\bar{a}\in T_r}$, $\{\gamma_{\bar{a}}\}_{\bar{a}\in T_r}$ and $\{\chi_{k_{\bar{a}}}\}_{\bar{a}\in T_r}$ in such a way that (1)–(4) hold (an event which has probability at least $1-\varepsilon$ by our inductive assumption).

Consider a vertex $\bar{a} \in \partial T_r = T_r \setminus T_{r-1}$. We want to explore the neighborhood of $k_{\bar{a}}$ in G_n. To this end, we note that for all $\bar{b} \in T_{r-1}$, the neighborhood of $k_{\bar{b}}$ is already determined by our conditioning on $\mathbf{k}^{(r)}(G_n)$, implying in particular that none of the edges sent out from $k_{\bar{a}}$ can hit a vertex $k \in K_{r-1}$, unless, of course, $\bar{a}$ is of type R, and k happens to be the parent of $k_{\bar{a}}$—in which case the edge between k and $k_{\bar{a}}$ is already present. To determine the children of type L of the vertex $k_{\bar{a}}$, we therefore have to condition on not hitting the set K_{r-1}. But apart from this, the process of determining the children of $k_{\bar{a}}$ is exactly the same as that of determining the children of the root k_0. Since $|K_r| \le N$, $k \ge \delta n$ for all $k \in K_r$, and $\chi_k \le C$ for all $k \in K_r$, we have that $\sum_{k\in K_r} \varphi_k \le C'/n$ for some $C' < \infty$, implying that conditioning on $k \notin K_{r-1} \subset K_r$ has only a negligible influence on the distribution of the children of $k_{\bar{a}}$. We may therefore proceed as in the proof of Lemma 3.5 to obtain a coupling between a sequence of i.i.d. random variables $x_{\bar{a},i}$ distributed uniformly in $[0, x_{\bar{a}}]$ and the children $k_{\bar{a},i}$ of $k_{\bar{a}}$ that are of type L. As before, we obtain that for n large enough, with probability at least $1-\varepsilon$, we have $|S_{k_{\bar{a},i}-1} - x_{\bar{a},i}| \le \varepsilon$.

Repeating this process for all $k_{\bar{a}} \in \partial K_r = K_r \setminus K_{r-1}$, we obtain a set of vertices L_{r+1} consisting of all children of type L with parents in ∂K_r. It is easy to see that with probability tending to one as $n \to \infty$, the set L_{r+1} has no intersection with K_r, so we will assume this for the rest of this proof.

Next we continue with the vertices of type R. Assume that we have already determined all children of type R for a certain subset $U_r \subset \partial K_r$, and denote the set children obtained so far by R_{r+1}. We decompose this set as $R_{r+1} = \bigcup_{i=1}^m R_{r+1}^{(i)}$, where $R_{r+1}^{(i)} = \{k \in R_{r+1} : X_{i,k} \in U_r\}$.

Consider a vertex $\bar{a} \in \partial K_r \setminus U_r$. Conditioning on the graph explored so far is again not difficult, and now amounts to two conditions:

(1) $X_{k,i} \neq k_{\bar{a}}$ if $k \in K_r \cup R_{r+1}^{(i)}$, since all the edges sent out from this set have already been determined.

(2) For $k \notin K_r \cup R_{r+1}^{(i)}$, the probability that $k_{\bar{a}}$ receives the ith edge from k is different from the probability given in (14), since the random variables $X_{k,i}$ has been probed before: we know that $X_{k,i} \notin K_{r-1}$ since otherwise k had sent out an edge to a vertex in K_{r-1}, which means that k would have been a child of type R in K_r. We also know that $X_{k,i} \notin U_r$, since otherwise $k \in R_{r+1}^{(i)}$. Instead of (14), we therefore have to use the modified probability

$$P_{k\to k_{\bar{a}}} = \varphi_{k_{\bar{a}}} \frac{1}{\tilde{S}_{k-1}},$$

where

$$\tilde{S}_{k-1} = \sum_{\substack{k' > k_{\bar{a}}: \\ k' \notin K_{r-1} \cup U_r}} \varphi_{k'}.$$

Since $\tilde{S}_{k-1} \leq S_{k-1} \leq \tilde{S}_{k-1} + C'/n$ by our inductive assumption, we can again refer to Lemma 3.1 to approximate $P_{k \to k_{\bar{a}}}$ by

$$\hat{P}_{k \to k_{\bar{a}}} = \frac{1}{nm} \frac{\chi k_{\bar{a}}}{2(1+u)} \frac{n}{k_{\bar{a}}} \left(\frac{k_{\bar{a}}}{k}\right)^{\chi}.$$

From here on the proof of our inductive claim is completely analog to the proof of Lemma 3.5. We leave it to the reader to fill in the (straightforward but slightly tedious) details. $\square$

3.6. *Estimates for the Pólya urn representation.* In this section we complete the work started in Section 3.1 by proving Lemmas 3.1 and 3.2.

PROOF OF LEMMA 3.1. Fix ε, and recall that

$$\chi = \frac{1+2u}{2+2u} \in \left[\frac{1}{2}, 1\right).$$

Writing S_k as

$$S_k = \prod_{j=k+1}^{n} (1 - \psi_j) = \exp\left(\sum_{j=k+1}^{n} \log(1 - \psi_j)\right),$$

we use the fact that if $0 < x < 1$, then $x \leq -\log(1-x) \leq x + x^2/(1-x)$ to bound

$$\left| E\left[\sum_{j=k+1}^{n} \log(1 - \psi_j)\right] + \sum_{j=k+1}^{n} E[\psi_j] \right| \leq \sum_{j=k+1}^{n} E\left[\frac{\psi_j^2}{1 - \psi_j}\right].$$

On the other hand, by Kolmogorov's inequality and the fact that

$$\operatorname{Var}(\log(1 - \psi_j)) \leq E\left[(\log(1 - \psi_k))^2\right] \leq E\left[\psi_j^2 (1 - \psi_j)^{-2}\right],$$

we have

$$\mathbf{P}\left(\max_{K \leq k \leq n} \left| \sum_{j=k+1}^{n} \log(1 - \psi_j) - E\left[\sum_{j=k+1}^{n} \log(1 - \psi_j)\right] \right| \geq \varepsilon\right)$$
$$\leq \frac{1}{\varepsilon^2} \sum_{j=K+1}^{n} E\left[\frac{\psi_j^2}{(1 - \psi_j)^2}\right].$$

We will use that for any $\beta_{a,b}$ distributed random variable ψ, we have

$$E[\psi] = \frac{a}{a+b}, \qquad E\left[\frac{\psi^2}{1 - \psi}\right] = \frac{a(a+1)}{(a+b)(b-1)}$$

and

$$E\left[\frac{\psi^2}{(1-\psi)^2}\right] = \frac{a(a+1)}{(b-2)(b-1)}.$$

Using these relations for $a = m + 2mu$ and $b = (2j-3)m + 2mu(j-1)$, we get

$$E(\psi_j) = \frac{m+2mu}{(2j-2)m+2jmu} = \frac{\chi}{j} + O\left(\frac{1}{j^2}\right), \tag{21}$$

$$E[\psi_j^2] \le E\left[\frac{\psi_j^2}{1-\psi_j}\right] = O\left(\frac{1}{j^2}\right) \quad \text{and} \quad E\left[\frac{\psi_j^2}{(1-\psi_j)^2}\right] = O\left(\frac{1}{j^2}\right). \tag{22}$$

Putting these bounds together, and observing that $\sum_{j=k+1}^{n} \frac{1}{j} = \log(n/k) + O(k^{-1})$, we get that there exists a constant $K(\varepsilon)$ not depending on n such that with probability at least $1-\varepsilon$, we have that

$$\left(\frac{k}{n}\right)^{\chi} e^{-\varepsilon} < S_k < \left(\frac{k}{n}\right)^{\chi} e^{\varepsilon} \qquad \text{for all } K(\varepsilon) \le k \le n.$$

For $k < K(\varepsilon)$, we bound $S_k \le S_K$ to conclude that with probability at least $1-\varepsilon$,

$$\left| S_k - \left(\frac{k}{n}\right)^{\chi} \right| = O\left(\left(\frac{K}{n}\right)^{\chi}\right).$$

The lemma now follows. $\square$

PROOF OF LEMMA 3.2. (i) Let $a = m + 2mu$, so that $\chi_k \sim \Gamma(a, 1)$. Then

$$\mathbf{P}(\chi_k \ge \log^2 k) \le E[e^{\chi_k/2}] e^{-(\log^2 k)/2} = 2^a k^{-(\log k)/2}.$$

Since the right-hand side is sumable, this implies the first statement of the lemma through the Borel–Cantelli lemma.

(ii) Let $b_k = (2k-3)m + 2mu(k-1) - 1$, and let $\chi_k' = \chi_k / b_k$. Then f_k can be defined by

$$\mathbf{P}(\psi_k \le f_k(x)) = \mathbf{P}(\chi_k' \le x/b_k).$$

In order to prove the second statement of the lemma, it is clearly enough to prove that for all sufficiently large k, we have

$$(1-\varepsilon)\frac{x}{b_k} \le f_k(x) \le \frac{x}{b_k} \qquad \text{for } x \le \log^2 k,$$

which in turn is equivalent to showing that

$$\mathbf{P}(\psi_k \le (1-\varepsilon)x) \le \mathbf{P}(\chi_k' \le x) \le \mathbf{P}(\psi_k \le x) \qquad \text{for } x \le \frac{\log^2 k}{b_k} \tag{23}$$

provided k is large enough.

We start by proving that

$$\Delta(x) := \mathbf{P}(\psi_k \le x) - \mathbf{P}(\chi'_k \le x) \ge 0.$$

To this end, we rewrite

$$\mathbf{P}(\psi_k \le x) = \frac{1}{Z_\beta}\int_0^x y^{a-1}(1-y)^b\,dy$$

and

$$\mathbf{P}(\chi'_k \le \lambda) = \frac{1}{Z_\gamma}\int_0^\lambda y^{a-1}e^{-by}\,dy,$$

where $a = m + 2mu$, $b = b_k$ and $Z_\gamma = \int_0^\infty y^{a-1}e^{-by}\,dy$ and $Z_\beta = \int_0^1 y^{a-1}(1-y)^b\,dy$ are the appropriate normalization factors. For $x \le 1$, we express $\Delta(x)$ as

$$\Delta(x) = \frac{1}{Z_\gamma}\int_0^x dy\,y^{a-1}e^{-by}\left(e^\delta \exp\left(-b\sum_{k=2}^\infty \frac{y^k}{k}\right) - 1\right),$$

where $e^\delta = Z_\gamma/Z_\beta$. Note that $\delta > 0$ by the fact that $(1-x) \le e^{-x}$. It is also easy to see that $\delta \to 0$ as $k \to \infty$; indeed, we have $\delta = O(b^{-1}) = O(k^{-1})$.

Consider the derivative

$$\frac{d\Delta(x)}{dx} = \frac{x^{a-1}e^{-bx}}{Z_\gamma}\left(e^\delta \exp\left(-b\sum_{k=2}^\infty \frac{x^k}{k}\right) - 1\right),$$

and let x_0 be the unique root, that is, let $x_0 \in (0, 1)$ be the solution of the equation

$$\delta = b\sum_{k=2}^\infty \frac{x_0^k}{k}.$$

Then $\Delta(x)$ is monotone increasing for $0 < x < x_0$ and monotone decreasing for all $x > x_0$. Together with the observation that $\Delta(x) > 0$ for all sufficiently small x, and $\Delta(x) \to 0$ as $x \to \infty$, we conclude that $\Delta(x) \ge 0$ for $0 \le x < \infty$. This proves that $\mathbf{P}(\chi'_k \le x) \le \mathbf{P}(\psi_k \le x))$ for all $x \ge 0$.

To prove the lower bound in (23), we will prove that

$$\tilde{\Delta}(x) = \mathbf{P}(\chi'_k \le x) - \mathbf{P}(\psi_k \le (1-\varepsilon)x) \ge 0 \qquad \text{if } x \le \frac{\varepsilon}{4} \le \frac{1}{8}.$$

We decompose the range of x into two regions, depending on whether $x \ge \frac{4a}{b\varepsilon}$ or $x \le \frac{4a}{\varepsilon b}$.

In the first region, we express $\tilde{\Delta}(x)$ as

$$\begin{aligned}\tilde{\Delta}(x) &= \mathbf{P}(\psi_k \ge (1-\varepsilon)x) - \mathbf{P}(\chi'_k \ge x)\\ &= \frac{e^\delta}{Z_\gamma}\int_{x(1-\varepsilon)}^1 dy\,y^{a-1}(1-y)^b - \frac{1}{Z_\gamma}\int_x^\infty dy\,y^{a-1}e^{-by}.\end{aligned}$$

We then bound

$$\begin{aligned}\int_x^\infty dy(2y)^{a-1}e^{-by} &\le \int_x^{2x} dyy^{a-1}e^{-by}\int_{2x}^\infty dyy^{a-1}e^{-by}\\ &\le \int_x^{2x} dyy^{a-1}e^{-by} + 2^{a-1}e^{-bx}\int_x^\infty dyy^{a-1}e^{-by}\end{aligned}$$

proving that

$$\begin{aligned}\int_x^\infty dy(2y)^{a-1}e^{-by} &\le \bigl(1-2^{a-1}e^{-bx}\bigr)^{-1}\int_x^{2x} dyy^{a-1}e^{-by}\\ &\le 2\int_x^{2x} dyy^{a-1}e^{-by},\end{aligned} \tag{24}$$

where we have used $bx \ge a\log 2$ in the last step.

On the other hand, using that $(1-y)^b \ge e^{-by(1+x)}$ if $y \le 2x \le 1/2$, we have that

$$\begin{aligned}e^\delta\int_{x(1-\varepsilon)}^1 dyy^{a-1}(1-y)^b &\ge \int_{x(1-\varepsilon)}^{2x(1-\varepsilon)} dyy^{a-1}e^{-by(1+x)}\\ &= \int_x^{2x} dyy^{a-1}(1-\varepsilon)^a e^{-by(1+x)(1-\varepsilon)}\\ &\ge (1-\varepsilon)^a e^{-2bx^2}e^{\varepsilon bx}\int_x^{2x} dyy^{a-1}e^{-by}\\ &\ge 2\int_x^{2x} dyy^{a-1}e^{-by}.\end{aligned}$$

Combined with (24), this proves that $\tilde{\Delta}(x) \ge 0$ if $\varepsilon bx \ge 4a$.

For $\varepsilon bx \le 4a$, we bound

$$\begin{aligned}\tilde{\Delta}(x) &= \frac{1}{Z_\gamma}\left(\int_0^x dyy^{a-1}e^{-by} - e^\delta\int_0^{x(1-\varepsilon)} dyy^{a-1}(1-y)^b\right)\\ &\ge \frac{1}{Z_\gamma}\left(\int_0^x dyy^{a-1}e^{-by} - e^\delta\int_0^{x(1-\varepsilon)} dyy^{a-1}e^{-by}\right)\\ &= \frac{1}{Z_\gamma}\left(\int_{(1-\varepsilon)x}^x dyy^{a-1}e^{-by} - \bigl(e^\delta-1\bigr)\int_0^{x(1-\varepsilon)} dyy^{a-1}e^{-by}\right)\\ &\ge \frac{1}{Z_\gamma}\bigl(\varepsilon x\bigl[(1-\varepsilon)x\bigr]^{a-1}e^{-bx} - \bigl(e^\delta-1\bigr)x^a\bigr)\\ &\ge \frac{x^a}{Z_\gamma}\bigl(\varepsilon 2^{1-a}e^{-4a/\varepsilon} - \bigl(e^\delta-1\bigr)\bigr).\end{aligned}$$

Since $\delta \to 0$ as $b \to \infty$, we see that the right-hand side becomes positive if $k \ge K$ for some $K < \infty$ that depends on a and ε (it grows exponentially in a/ε). $\square$

4. Approximating coupling for the independent and the conditional models. In this section we prove that the sequential and the independent model have the same weak limit. To this end we construct a coupling between the two models such with probability tending to 1, the balls around a randomly chosen vertex in $\{1, \dots, n\}$ are identical in both models. This will imply that both models have the same weak local limit.

We only give full details for the coupling between the independent and the sequential model. The approximating coupling between the conditional and the sequential model is very similar, and the proof that it works is identical.

We construct the coupling inductively as follows: let $V = 1, 2, \dots$ be the vertices of the preferential attachment graph. For $1 \neq n \in V$ and $i = 1, \dots, m$ let $e_n^i < n$ and $f_n^i < n$ be the ith vertex that n is connected to in, respectively, the sequential and the independent models. We use the symbol $\mathbf{e}_n$ to denote the vector $\{e_n^i\}_{1 \le i \le m}$, and the symbol $\mathbf{f}_n$ to denote the vector $\{f_n^i\}_{i=1}^m$.

By construction, $e_2^i = f_2^i = 1$ for all i. Once we know $\mathbf{e}_l$ and $\mathbf{f}_l$ for every $l < n$, we determine $\mathbf{e}_n$ and $\mathbf{f}_n$ as follows: let D_1 be the distribution of $\mathbf{e}_n$, based on the sequential rule and conditioned on $\{\mathbf{e}_l\}_{l<n}$, and let D_2 be the distribution of $\mathbf{f}_n$ based on the independent rule and conditioned on $\{\mathbf{f}_l\}_{l<n}$. Let D be an (arbitrarily chosen) coupling of D_1 and D_2 that minimizes the total variation distance. Then we choose $\mathbf{e}_n$ and $\mathbf{f}_n$ according to D.

Our goal is to prove the following proposition:

PROPOSITION 4.1. *Let (G_n) and (G'_n) be the sequence of preferential attachment graphs in the sequential and the conditional model, respectively, coupled as above. Let $\varepsilon > 0$ and let r be an arbitrary positive integer. Then there exists an integer n_0 such that for $n \ge n_0$, with probability at least $1 - \varepsilon$, a uniformly chosen random vertex $k_0 \in \{1, \dots, n\}$ has the same r-neighborhood in G_n and G'_n.*

The proof of the proposition relies on following two lemmas, to be proven in Sections 4.2 and 4.3, respectively.

LEMMA 4.2. *Consider the coupling defined above, and fix $k \ge 2$. For $n > k$, let $A_n = A_n^{(k)}$ be the event that there exists an $i \in \{1, \dots, m\}$ such that $e_n^i = k \neq f_n^i$ or $e_n^i \neq k = f_n^i$. Then*

$$\mathbf{P}\left(A_n \,\middle|\, \bigcap_{h=k+1}^{n-1} A_h^c, d_{n-1}(k)\right) = O\left(\frac{d_{n-1}(k)}{n^2}\right). \tag{25}$$

Note that under the conditioning, $d_{n-1}(k)$ is the same in both models.

LEMMA 4.3. *For the sequential preferential attachment model, for every n and k such that $n > k$, let $d_n(k)$ be the degree of vertex k when the graph contains* *n vertices. Then*

$$E\big[d_n(k)\big] = m\left[1 + \frac{\chi}{1-\chi}\left(\left(\frac{n}{k}\right)^{1-\chi} - 1\right)\right] + O\left(\frac{n^{1-\chi}}{k^{2-\chi}}\right), \tag{26}$$

where the constant implicit in the O-symbol depends on m and u.

4.1. *Proof of Proposition* 4.1. Fix ε and r, let $B_r(k)$ and $B_r(k)'$ be the ball of radius r about k in G_n and G'_n, respectively, and let B be the set of vertices $k \in \{1, \dots, n\}$ for which $B_r(k) \neq B_r(k)'$. Then the probability that a uniformly chosen vertex in $\{1, \dots, n\}$ is in B is just $1/n$ times the expected size of B. We thus have to show that

$$E\big[|B|\big] \leq \varepsilon n.$$

In a preliminary step note that $B_r(k) = B_r(k)'$ unless there exists a vertex $k' \in B_r(k)$ such that $e^i_{n'} = k' \neq f^i_{n'}$ or $e^i_{n'} \neq k' = f^i_{n'}$ for some $i = 1, \dots, m$ and some $n' > k'$.

To prove this fact, let us consider the event $A^{(k)} = \bigcup_{n>k} A^{(k)}_n$. It is easy to see that this event is the event that at least one of the edges received by k is different in (G_n) and (G'_n). Using this fact, one easily shows that the ball of radius 1 around a vertex k must be identical in G_n and G'_n unless $A^{(k')}$ happens for at least one vertex k' in the 1-neighborhood of k in G_n. By induction, this implies that $B_r(k) = B_r(k)'$ unless there exists a vertex $k' \in B_r(k)$ such that the event $A^{(k')}$ happens, which is what we claimed in the previous paragraph.

Next we note that by Proposition 3.6, there exist $\delta > 0$ and $N < \infty$ such that with probability at least $1 - \varepsilon/2$, a random vertex $k \in \{1, \dots, n\}$ obey the two following two conditions:

(1) the ball of radius $2r$ around k in the sequential graph G_n contains no more than N vertices;

(2) the oldest vertex (the vertex with the smallest index) in this ball is no older than δn.

If we denote the set of vertices satisfying these two conditions by W, we thus have that

$$E\big[|W|\big] \geq \left(1 - \frac{\varepsilon}{2}\right)n.$$

As a consequence, it will be enough to show that

$$E\big[|W \cap B|\big] \leq \frac{\varepsilon}{2} n.$$

If $k \in W \cap B$, there must be a vertex $k' \in B_r(k)$ such that the event $A^{(k)}$ happens. But $k' \in B_r(k)$ if and only if $k \in B_r(k')$, and since $B_r(k') \subset B_{2r}(k)$, we must further have that $|B_r(k')| \le N$ and $k' \ge \delta n$. As a consequence,

$$\begin{aligned}
|W \cap B| &= \sum_{k\in W} \mathbf{I}(k \in B) \le \sum_{k\in W} \sum_{k'\in B_r(k)} \mathbf{I}\bigl(A^{(k')}\bigr) \\
&= \sum_{k'} \mathbf{I}\bigl(A^{(k')}\bigr) \sum_{k\in B_r(k')} I(k \in W) \\
&\le N \sum_{k'=\delta n}^{n} \mathbf{I}\bigl(A^{(k')}\bigr),
\end{aligned}$$

where we used the symbol $\mathbf{I}(A)$ to denote the indicator function of the event A.

Finally by Lemmas 4.2 and 4.3,

$$P\bigl(A^{(k)}\bigr) \le O(1) \sum_{n>k} \frac{1}{n^2}\left(\frac{n}{k}\right)^{1-\chi} = O\left(\frac{1}{k}\right).$$

As a consequence we can find a constant C such that

$$E\bigl[|W \cap B|\bigr] \le N \sum_{k'=n\delta}^{n} \frac{C}{k'} \le CN/\delta.$$

For n large enough, the right-hand side is smaller than $\frac{\varepsilon}{2}n$, which is the bound we had to establish.

4.2. *Proof of Lemma* 4.2. Let us the shorthand d for the degree d_{n-1}^k. In the independent model the probability of having r connections to k and $h = m - r$ connections to other vertices in $\{1, \dots, n-1\}$ is

$$\binom{m}{r} p^r (1-p)^h \qquad \text{with } p = \frac{\alpha}{n-1} + \frac{(1-\alpha)d}{2m(n-2)},$$

while in the sequential model it is

$$\binom{m}{r} \prod_{l=0}^{r-1} p_l \prod_{l=r}^{m-1} (1 - p_l)$$

with

$$p_l = p_l(r) = \begin{cases} \dfrac{2m\alpha + (1-\alpha)(d+l)}{2m(n-2) + 2m\alpha + (1-\alpha)l}, & \text{if } l < r, \\ \dfrac{2m\alpha + (1-\alpha)(d+r)}{2m(n-2) + 2m\alpha + (1-\alpha)l}, & \text{if } l \ge r. \end{cases}$$

[Here we used exchangeability and (6).]

As a consequence, the probability in (25) is bounded by a constant times

$$\max_{r=0,\dots,m} \left| \left[\prod_{l=0}^{r-1} p_l \prod_{l=r}^{m-1} (1 - p_l) \right] - p^r (1-p)^h \right|. \tag{27}$$

Telescoping the difference, we bound (27) by

$$\max_{r=0,\dots,m}\left(\sum_{l=0}^{r-1} p^l |p_l - p| \prod_{l'=l+1}^{r-1} p_{l'}\right.$$
$$\left. + \sum_{l=r}^{m-1} p^r (1-p)^{l-r} \big|(1-p)-(1-p_l)\big| \prod_{l'=l+1}^{m-1} (1-p_{l'})\right)$$
$$\leq \max_{r=0,\dots,m}\left(\tilde{p}^{r-1}\sum_{l=0}^{r-1}|p_l - p| + p^r \sum_{l=r}^{m-1}|p-p_l|\right),$$

where $\tilde{p} = \max\{p, p_1, \dots, p_m\} = O(d/n)$. We now distinguish three cases:

(i) if $r \geq 2$, we use the fact that $p - p_l = O(1/n)$ to get a bound of order $O(\tilde{p}/n) = O(d/n^2)$ for both sums;

(ii) if $r = 1$, we use the fact that the first sum is equal to $|p_0 - p| = O(1/n^2)$, while the second can be bounded by $O(\tilde{p}/n) = O(d/n^2)$ as before;

(iii) if $r = 0$, we use that fact that

$$\begin{aligned} p_l(0) &= \frac{2m\alpha + (1-\alpha)d}{2m(n-2) + 2m\alpha + (1-\alpha)l} \\ &= \frac{2m\alpha}{2m(n-1)}\big(1 + O(n^{-1})\big) + \frac{(1-\alpha)d}{2m(n-2)}\big(1 + O(n^{-1})\big) \\ &= p + O\big(d/n^2\big) \end{aligned}$$

to show that for $r = 0$, all terms in the sum $\sum_{l=0}^{m-1} |p - p_l|$ are of order $O(d/n^2)$.

This completes the proof of the lemma.

4.3. *Proof of Lemma* 4.3. As before, we use $\varphi_k^{(n)}$ for

$$\varphi_k^{(n)} = \psi_k \prod_{i=k+1}^{n} (1-\psi_i).$$

By construction,

$$d_n(k) = m + \sum_{t=(k-1)m+1}^{(n-1)m} \mathcal{U}_t, \tag{28}$$

where the variables $\{\mathcal{U}_t\}$ are defined as follows: let $\{\hat{U}_t\}_{t=1}^{\infty}$ be i.i.d. $U[0,1]$ variables, independent of the φ_k's. Then $\mathcal{U}_t = \mathbf{1}_{\hat{U}_t < \varphi_k^{(\lceil t/m \rceil)}}$. Note that conditioned on $\{\varphi_k^{(j)}\}_{j \geq k}$, $\{\mathcal{U}_t\}$'s are independent, each being Bernoulli $\varphi_k^{(\lceil t/m \rceil)}$.

Let $\mathcal{F}$ be the σ-algebra generated by $\{\psi_h\}_{h=1}^{\infty}$. Then

$$E\bigl(d_n(k)\mid\mathcal{F}\bigr)=m+m\sum_{\ell=k}^{n-1}\varphi_k^{(\ell)}. \tag{29}$$

By (21),

$$\frac{\chi}{k}\le E(\psi_k)\le\frac{\chi}{k-1}, \tag{30}$$

which in turn implies that

$$\begin{aligned}E[\varphi_k^{(\ell)}]&=E[\psi_k]\prod_{i=k+1}^{\ell}\bigl(1-E[\psi_i]\bigr)\le\frac{\chi}{k-1}\prod_{i=k+1}^{\ell}\Bigl(1-\frac{\chi}{i}\Bigr)\\&\le\frac{\chi}{k-1}\exp\Biggl(-\chi\sum_{i=k+1}^{\ell}\frac{1}{i}\Biggr)\le\frac{\chi}{k-1}\exp\Bigl(-\chi\log\Bigl(\frac{\ell+1}{k+1}\Bigr)\Bigr)\\&=\frac{\chi}{k-1}\Bigl(\frac{k+1}{\ell+1}\Bigr)^{\chi},\end{aligned}$$

implying that

$$\begin{aligned}E\bigl[d_n(k)\bigr]&\le m+m\chi\frac{(k+1)^{\chi}}{k-1}\sum_{\ell=k}^{n-1}\Bigl(\frac{1}{\ell+1}\Bigr)^{\chi}\\&\le m+m\chi\frac{(k+1)^{\chi}}{k-1}\int_{k-1}^{n-1}dx\Bigl(\frac{1}{x+1}\Bigr)^{\chi}\\&=m+m\frac{\chi}{1-\chi}\frac{(k+1)^{\chi}}{k-1}\bigl(n^{1-\chi}-k^{1-\chi}\bigr)\\&\le m+m\frac{\chi}{1-\chi}\frac{k+1}{k-1}\Bigl(\Bigl(\frac{n}{k}\Bigr)^{1-\chi}-1\Bigr)\\&\le m+m\frac{\chi}{1-\chi}\Bigl(\Bigl(\frac{n}{k}\Bigr)^{1-\chi}-1\Bigr)\Bigl(1+\frac{4}{k}\Bigr).\end{aligned} \tag{31}$$

On the other hand, again by (30),

$$\begin{aligned}E[\varphi_k^{(\ell)}]&\ge\frac{\chi}{k}\prod_{i=k+1}^{\ell}\Bigl(1-\frac{\chi}{i-1}\Bigr)\ge\frac{\chi}{k}\prod_{i=k+1}^{\ell}\Bigl(1-\frac{1}{i-1}\Bigr)^{\chi}\\&=\frac{\chi}{k}\prod_{i=k+1}^{\ell}\Bigl(\frac{i-2}{i-1}\Bigr)^{\chi}=\frac{\chi}{k}\Bigl(\frac{k-1}{\ell-1}\Bigr)^{\chi}\end{aligned}$$

implying that

$$
\begin{aligned}
E\big[d_n(k)\big] &\geq m + m\chi \frac{(k-1)^\chi}{k} \sum_{\ell=k}^{n-1} \left(\frac{1}{\ell-1}\right)^\chi \\
&\geq m + m\chi \frac{(k-1)^\chi}{k} \int_k^n dx \left(\frac{1}{x-1}\right)^\chi \\
&= m + m\frac{\chi}{1-\chi} \frac{(k-1)^\chi}{k} \big((n-1)^{1-\chi} - (k-1)^{1-\chi}\big) \qquad (32) \\
&= m + m\frac{\chi}{1-\chi} \frac{k-1}{k} \left(\left(\frac{n-1}{k-1}\right)^{1-\chi} - 1\right) \\
&\geq m + m\frac{\chi}{1-\chi} \left(\left(\frac{n}{k}\right)^{1-\chi} - 1\right)\left(1 - \frac{1}{k}\right).
\end{aligned}
$$

5. Applications.

5.1. *Degree distribution of an early vertex.* In this section, we will show that for $n \gg k \gg 1$, $d_n(k)$ grows like $(\frac{n}{k})^{1-\chi} = (\frac{n}{k})^{\psi/(\psi+1)}$. To give the precise statement, we need some definition. To this end, let us consider the random variables

$$M_k^{(\ell)} = \prod_{j=k+1}^{\ell} \frac{1-\psi_j}{1-E[\psi_j]}.$$

The bounds (21) and (22) imply that the second moment of $M_k^{(\ell)}$ is bounded uniformly in ℓ, so by the martingale convergence theorem, $M_k^{(\ell)}$ converges both a.s. and in L^2. Since $1 - E[\psi_j] = (\frac{j-1}{j})^\chi + O(j^{-2})$, this also implies that the limit

$$F_k = \lim_{\ell\to\infty} \prod_{j=k+1}^{\ell} (1-\psi_j)\left(\frac{j}{j-1}\right)^\chi = \lim_{\ell\to\infty} \left(\frac{\ell}{k}\right)^\chi \prod_{j=k+1}^{\ell} (1-\psi_j) \qquad (33)$$

exists a.s. and in L^2. In the following lemma, $O_P(k^{-1/2})$ stand for a random variable A such that $Ak^{1/2}$ is bounded in probability.

LEMMA 5.1. *Consider the sequential model for some α and m, and let F_k be as above. Then*

$$\frac{d_n(k)}{n^{1-\chi}} \to \frac{m}{1-\chi} k^\chi \psi_k F_k \qquad \text{as } n \to \infty, \qquad (34)$$

both in expectation and in distribution. Furthermore,

$$F_k > 0 \qquad \text{a.s. for all } k \geq 1, \qquad \log F_k = O_P\big(k^{-1/2}\big)$$

and

$$E[F_k] = 1 + O(k^{-1}),$$

implying in particular that

$$\lim_{n\to\infty} \frac{E[d_n(k)]}{n^{1-\chi}} = \frac{m\chi}{1-\chi}\frac{1}{k^{1-\chi}}\bigl(1 + O(k^{-1})\bigr).$$

REMARK. Note that (34) holds also for the independent and the conditional models. The reason is that by the approximating coupling, the total variation distance between the degree distribution of vertex number k in the sequential model and that of vertex number k in the independent (or conditional) model goes to 0 as k goes to infinity, and the convergence is uniform in n (the size of the graph).

PROOF OF LEMMA 5.1. We first consider the conditional expectation $E[d_n(k) \mid \mathcal{F}]$, where, as before, $\mathcal{F}$ is the σ-algebra generated by $\{\psi_h\}_{h=1}^{\infty}$. Fix ε, and let K be such that for $\ell \geq K$,

$$\left\| F_k - \left(\frac{\ell}{k}\right)^{\chi} \prod_{j=k+1}^{\ell} (1-\psi_j) \right\|_2 \leq \varepsilon.$$

Bounding

$$\left\| E\bigl[d_n(k) \mid \mathcal{F}\bigr] - \sum_{\ell=K}^{n-1} m\varphi_k^{(\ell)} \right\|_2 \leq mK$$

we then approximate

$$\begin{aligned}\sum_{\ell=K}^{n-1} m\varphi_k^{(\ell)} &= m\psi_k \sum_{\ell=K}^{n-1} \prod_{j=k+1}^{\ell} (1-\psi_j) = m\psi_k \sum_{\ell=K}^{n-1} \left(\frac{k}{\ell}\right)^{\chi} \bigl(F_k + O(\varepsilon)\bigr) \\ &= n^{1-\chi}\left(\frac{m}{1-\chi} k^{\chi}\psi_k F_k + O(\varepsilon)\right),\end{aligned}$$

where the errors $O(\varepsilon)$ stand for errors in L^2. We thus have show that as $n \to \infty$,

$$\frac{1}{n^{1-\chi}} E\bigl[d_n(k) \mid \mathcal{F}\bigr] \to \frac{m}{1-\chi} k^{\chi}\psi_k F_k$$

in L^2. Taking expectations on both sides, we obtain that (34) holds in expectation.

To prove convergence in distribution, it is clearly enough to show that $E[d_n(k) \mid \mathcal{F}] - d_n(k) \to 0$ in probability. But this follows by an easy second moment estimate and the observation that

$$E\bigl[d_n(k)^2 \mid \mathcal{F}\bigr] \leq E\bigl[d_n(k) \mid \mathcal{F}\bigr]^2 + E\bigl[d_n(k) \mid \mathcal{F}\bigr].$$

Next we observe that the bounds established in Section 3.6 imply that there is a constant $C < \infty$ such that for $k \geq 2$,

$$\left|\log M_k^{(\ell)}\right| \leq \varepsilon + \frac{C}{k}$$

with probability at least $1 - \frac{C}{\varepsilon^2 k}$. Since these bounds are uniform in ℓ, they carry over to the limit, and imply both that a.s. $F_k > 0$ for all fixed $k \geq 2$, and that $\log F_k = O_P(k^{-1/2})$ as $k \to \infty$. To prove that a.s. $F_1 > 0$, we note that F_1/F_2 is proportional to $1 - \psi_2$. The bound $E[F_k] = 1 + O(k^{-1})$ finally follows from the fact that $E[M_k^{(\ell)}] = 1$ and the observation that $1 - E[\psi_j] = (\frac{j-1}{j})^{\chi} + O(j^{-2})$. $\square$

5.2. *Degree distribution.* By Theorem 2.2 and Corollary 2.3, the limiting degree distribution of the preferential attachment graph G_n is exactly the degree distribution of the root of the Pólya-point graph. As we will see, this allows us to explicitly calculate the limiting degree distribution of the preferential attachment graph. In a similar way, it also allows us to calculate the limiting degree distribution of a vertex chosen at random from the vertices that receive an edge from a uniformly random vertex v_0 in G_n. We summarize the results in the following lemma.

LEMMA 5.2. *Let v_0 be a uniformly chosen vertex in G_n, let D be the degree of v_0 and let D' be the degree of a vertex chosen uniformly at random from the m vertices which received an edge from v_0. In the limit $n \to \infty$, the distribution of D and D' for all three versions of the preferential attachment graph converge to*

$$\mathbf{P}(D = m + k) = \frac{\psi + 1}{\psi} \frac{\Gamma(a + 1/\psi + 1)}{\Gamma(a)} \frac{\Gamma(k + a)}{\Gamma(a + 1/\psi + k + 2)}$$

and

$$\mathbf{P}(D' = m + 1 + k) = \frac{\psi + 1}{\psi^2} \frac{\Gamma(a + 1/\psi + 1)}{\Gamma(a + 1)} \frac{(k + 1)\Gamma(k + a + 1)}{\Gamma(a + 1/\psi + k + 3)},$$

where $a = m + 2mu$. As $k \to \infty$, this gives

$$\mathbf{P}(D = m + k) = Ck^{-2-1/\psi}\left(1 + O\left(k^{-1}\right)\right)$$

and

$$\mathbf{P}(D' = m + 1 + k) = \tilde{C}k^{-1-1/\psi}\left(1 + O\left(k^{-1}\right)\right)$$

for some constants C and $\tilde{C}$ depending on m and α.

Note that for $\alpha = 0$, the statements of the lemma reduce to

$$\mathbf{P}(D = m + k) = \frac{2m(m + 1)}{(m + k)(m + k + 1)(m + k + 2)}$$

and

$$\mathbf{P}\bigl(D' = m+1+k\bigr) = \frac{2(m+1)(k+1)}{(m+k+1)(m+k+2)(m+k+3)}.$$

PROOF OF LEMMA 5.2. First we condition on the position x_0 of the root of the Pólya graph. Let D be the degree of the root. D conditioned on x_0 is m plus a Poisson variable with parameter

$$\frac{\gamma}{x_0^{\psi}} \int_{x_0}^{1} \psi x^{\psi-1}\,dx = \gamma \frac{1-x_0^{\psi}}{x_0^{\psi}},$$

where γ is a Gamma variable with parameters $a = m + 2mu$ and 1. Let

$$\kappa = \kappa(x_0) = \frac{1-x_0^{\psi}}{x_0^{\psi}}.$$

Then

$$\begin{aligned}
\mathbf{P}(D = m+k \mid x_0) &= \frac{\Gamma(k+a)}{k!\Gamma(a)} \frac{\kappa^k}{(\kappa+1)^{k+a}} \\
&= \frac{\Gamma(k+a)}{k!\Gamma(a)} \frac{(1-x_0^{\psi})^k}{x_0^{k\psi}} \bigl(x_0^{\psi}\bigr)^{k+a} \qquad (35)\\
&= \frac{\Gamma(k+a)}{k!\Gamma(a)} \bigl(1-x_0^{\psi}\bigr)^k x_0^{a\psi}
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{P}(D = m+k) &= (\psi+1) \int_0^1 \mathbf{P}(D = m+k \mid x_0 = x) x^{\psi}\,dx \\
&= (\psi+1) \frac{\Gamma(k+a)}{k!\Gamma(a)} \int_0^1 \bigl(1-x^{\psi}\bigr)^k x^{(a+1)\psi}\,dx \\
&= \frac{\psi+1}{\psi} \frac{\Gamma(k+a)}{k!\Gamma(a)} \int_0^1 (1-y)^k y^{a+1/\psi}\,dy \\
&= \frac{\psi+1}{\psi} \frac{\Gamma(k+a)}{\Gamma(a)} \prod_{i=1}^{k+1} \frac{1}{a+1/\psi+i} \\
&= \frac{\psi+1}{\psi} \frac{\Gamma(k+a)}{\Gamma(a)} \frac{\Gamma(a+1/\psi+1)}{\Gamma(a+1/\psi+k+2)}.
\end{aligned}$$

To calculate the distribution of D', we chose y_0 uniformly at random from $[0, x_0]$. Conditioned on y_0, the limiting degree D' is equal to $m+1$ plus a Poisson variable with parameter

$$\frac{\gamma'}{y_0^\psi}\int_{y_0}^1 \psi x^{\psi-1}\,dx = \gamma'\frac{1-y_0^\psi}{y_0^\psi},$$

where γ' is a Gamma variable with parameters $a+1$ and 1. Continuing as before, this gives

$$\mathbf{P}\big(D' = m+1+k \mid y_0 = y\big) = \frac{\Gamma(k+a+1)}{k!\Gamma(a+1)}\big(1-y^\psi\big)^k y^{(a+1)\psi} \tag{36}$$

$$\begin{aligned}\mathbf{P}\big(D' = m+1+k\big) &= (\psi+1)\int_0^1 dx_0 x_0^\psi \frac{1}{x_0}\int_0^{x_0} dy\mathbf{P}\big(D' = m+1+k \mid y_0 = y\big)\\ &= (\psi+1)\frac{\Gamma(k+a+1)}{k!\Gamma(a+1)}\int_0^1 dx x^{\psi-1}\int_0^x dy\big(1-y^\psi\big)^k y^{(a+1)\psi}\\ &= \frac{\psi+1}{\psi^2}\frac{\Gamma(k+a+1)}{k!\Gamma(a+1)}\int_0^1 du\int_0^u dv(1-v)^k v^{a+1/\psi}.\end{aligned}$$

Exchanging the integral over u and v we obtain

$$\begin{aligned}\mathbf{P}\big(D' = m+1+k\big) &= \frac{\psi+1}{\psi^2}\frac{\Gamma(k+a+1)}{k!\Gamma(a+1)}\int_0^1 dv(1-v)^k v^{a+1/\psi}\int_v^1 du\\ &= \frac{\psi+1}{\psi^2}\frac{\Gamma(k+a+1)}{k!\Gamma(a+1)}\int_0^1 dv(1-v)^{k+1} v^{a+1/\psi}\\ &= \frac{\psi+1}{\psi^2}\frac{(k+1)\Gamma(k+a+1)}{\Gamma(a+1)}\frac{\Gamma(a+1/\psi+1)}{\Gamma(a+1/\psi+k+3)}.\end{aligned}$$

The asymptotic behavior as $k\to\infty$ follows from the well-known asymptotic behavior of the Gamma function. $\square$

5.3. *Joint degree distributions.* We can use the same calculation in order to determine the joint distribution of the degree of the root of the preferential attachment graph with a vertex chosen uniformly among the m vertices that receive an edge from the root.

LEMMA 5.3. *Let v_0 be a uniformly chosen vertex in G_n, let D be the degree of v_0 and let D' be the degree of a vertex chosen uniformly at random from the m vertices which received an edge from v_0. In the limit $n\to\infty$, the joint distribution of D and D' for all three versions of the preferential attachment graph converges to*

$$\begin{aligned}&\mathbf{P}\big(D' = m+1+k, D = m+j\big)\\ &\quad= \frac{\psi+1}{\psi^2}\frac{\Gamma(k+a+1)}{k!\Gamma(a+1)}\frac{\Gamma(j+a)}{j!\Gamma(a)}\int_0^1 dv(1-v)^k v^{a+1/\psi}\int_v^1 du(1-u)^j u^a,\end{aligned}$$

where $a = m + 2mu$. As $k \to \infty$ while j is fixed, this gives

$$\mathbf{P}(D' = m + 1 + k \mid D = m + j) = C_j k^{-1-1/\psi}\left(1 + O\left(\frac{1}{k}\right)\right),$$

where C_j is a constant depending on j, m and α, while for k fixed and $j \to \infty$, we have

$$\mathbf{P}(D = m + j \mid D' = m + 1 + k) = \tilde{C}_k j^{-a-3-1/\psi}\left(1 + O\left(\left(\frac{1}{j}\right)\right)\right),$$

where $\tilde{C}_k$ is a constant depending on k, m and α.

Note that the conditioning on D does not change the power law for the degree distribution of D', while the conditioning on D' leads to a much faster falloff for the degree distribution of D. Intuitively, this can be explained by the fact that earlier vertices tend to have higher degree. Conditioning on the degree D' to be a fixed number therefore makes it more likely that at least one of the m vertices receiving an edge from v_0 was born late, which in turn makes it more likely that v_0 was born late. This in turn makes it much less likely that the root v_0 has very high degrees, leading to a faster decay at infinity. This effect does not happen for the distribution of D' conditioned on D, since the vertices receiving edges from the root are born *before* the root. Note the fact that the exponent of the power law of the distribution of D conditioned on D' depends (through a) on m. Heuristically, this seemingly surprising result follows from the fact that the distribution of the degree of the vertex at time k is (in the limit) a discretized Gamma distribution with parameter a (i.e., the probability of being equal k is proportional to $e^{-k/\lambda} \cdot k^a$. λ here is basically an appropriate power of n/k). Note that with this distribution, when λ is relatively large the probability of the degree being small is approximately λ^{-a}. This means that when D' is small, the probability that k is small (i.e., n/k is large) is as small as $(n/k)^a$. But for D to be big, k needs to be small (up to an exponential tail). This is the intuitive explanation for the parameter a comes into the exponent of the joint distribution.

PROOF OF LEMMA 5.3. Let x_0 be the location of the root in the Pólya-point graph, and let y_0 be the location of a vertex chosen uniformly at random from the m vertices of type L connected to the root. Then

$$\begin{aligned}
&\mathbf{P}(D' = k + m + 1, D = j + m) \\
&\quad = (\psi + 1)\int_0^1 dx x^{\psi-1}\int_0^x dy \mathbf{P}(D' = k + m \mid y_0 = y) \\
&\qquad \times \mathbf{P}(D = j + m \mid x_0 = x).
\end{aligned}$$

Using (35) and (36), we can write this explicitly a

$$\begin{aligned}
&\mathbf{P}\big(D' = k+m+1, D = j+m\big)\\
&\quad = (\psi+1)\frac{\Gamma(k+a+1)}{k!\Gamma(a+1)}\frac{\Gamma(j+a)}{j!\Gamma(a)}\int_0^1 dx\big(1-x^{\psi}\big)^j x^{(a+1)\psi-1}\\
&\qquad \times \int_0^x dy\big(1-y^{\psi}\big)^k y^{(a+1)\psi}\\
&\quad = \frac{\psi+1}{\psi^2}\frac{\Gamma(k+a+1)}{k!\Gamma(a+1)}\frac{\Gamma(j+a)}{j!\Gamma(a)}\int_0^1 du(1-u)^j u^a \int_0^u dv(1-v)^k v^{a+1/\psi}\\
&\quad = \frac{\psi+1}{\psi^2}\frac{\Gamma(k+a+1)}{k!\Gamma(a+1)}\frac{\Gamma(j+a)}{j!\Gamma(a)}\int_0^1 dv(1-v)^k v^{a+1/\psi}\int_v^1 du(1-u)^j u^a.
\end{aligned}$$

We want to approximate the double integral by a product of integrals. Clearly

$$\begin{aligned}
&\int_0^1 dv(1-v)^k v^{a+1/\psi}\int_v^1 du(1-u)^j u^a\\
&\qquad \le \int_0^1 dv(1-v)^k v^{a+1/\psi}\int_0^1 du(1-u)^j u^a\\
&\qquad = k!j!\frac{\Gamma(a+1/\psi+1)}{\Gamma(a+1/\psi+k+2)}\frac{\Gamma(a+1)}{\Gamma(a+j+2)} := Z.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\int_0^1 dv(1-v)^k v^{a+1/\psi}\int_0^v du(1-u)^j u^a\\
&\qquad \le \int_0^1 dv(1-v)^k v^{a+1/\psi}\int_0^v du u^a\\
&\qquad = \frac{1}{a+1}k!\frac{\Gamma(2a+1/\psi+2)}{\Gamma(2a+1/\psi+k+3)}\\
&\qquad = \frac{\Gamma(2a+1/\psi+2)}{\Gamma(a+1/\psi+1)\Gamma(a+2)}\frac{\Gamma(a+1/\psi+k+2)}{\Gamma(2a+1/\psi+k+3)}\frac{\Gamma(a+j+2)}{j!}Z\\
&\qquad = O\bigg(\Big(\frac{j}{k}\Big)^{a+1}\bigg)Z,
\end{aligned}$$

implying that

$$\begin{aligned}
&\mathbf{P}\big(D' = k+m+1 \mid D = j+m\big)\\
&\qquad = \frac{1}{\psi}\frac{\Gamma(k+a+1)}{\Gamma(a+1/\psi+k+2)}\frac{\Gamma(a+1/\psi+j+2)}{\Gamma(a+j+2)}\bigg(1+O\bigg(\Big(\frac{j}{k}\Big)^{a+1}\bigg)\bigg).
\end{aligned} \tag{37}$$

A similar calculation gives

$$\begin{aligned}\int_0^1 du(1-u)^j u^a \int_0^u dv(1-v)^k v^{a+1/\psi} \\ = \frac{j!}{a+1/\psi+1}\frac{\Gamma(2a+1/\psi+2)}{\Gamma(2a+1/\psi+j+3)}\left(1+O\left(\frac{k}{j}\right)\right),\end{aligned}$$

which in turn implies that for fixed k, as j goes to infinity, we get

$$\begin{aligned}&\mathbf{P}(D=j+m \mid D'=k+m+1) \\ &\quad = \frac{\Gamma(2a+1/\psi+2)}{\Gamma(a)\Gamma(a+1/\psi+2)}\frac{\Gamma(j+a)}{\Gamma(2a+1/\psi+j+3)} \\ &\qquad \times \frac{\Gamma(a+1/\psi+k+3)}{\Gamma(k+2)}\left(1+O\left(\frac{k}{j}\right)\right).\end{aligned} \tag{38}$$

The statements of the lemma describing the decay of (37) and (38) as (resp.) $k \to \infty$ and $j \to \infty$ follow from the well-known asymptotics of the Γ-function. $\square$

5.4. *Subgraph frequencies.*

5.4.1. *Proof of Lemma* 2.4. Let F be a finite graph with vertex set $V(F) = \{v_1, v_2, \dots, v_k\}$. As in Section 2.5, let $\mathrm{inj}(F, \mathbf{n}; G_n)$ be the number of injective maps Φ from $V(F)$ into $V(G_n)$ that are homomorphisms and preserve the degrees. In a similar way, given two rooted graphs (F, v) and (G, x), let $\widehat{\mathrm{inj}}((F, v), \mathbf{n}; (G_n, x))$ be the number of injective maps Φ from $V(F)$ into $V(G_n)$ that are homomorphisms, preserve the degrees and map v into x. Then $\mathrm{inj}(F, \mathbf{n}; G_n)$ can be reexpressed as

$$\mathrm{inj}(F, \mathbf{n}; G_n) = \sum_{x_1 \in V(G_n)} \widehat{\mathrm{inj}}\big((F, v_1), \mathbf{n}; (G_n, x_1)\big).$$

Since the diameter of (F, v_1) is at most k, its image under a homomorphism Φ has diameter at most k as well, which in turn implies that

$$\frac{1}{n}\mathrm{inj}(F, \mathbf{n}; G_n) = \frac{1}{n}\sum_{x_1 \in V(G_n)} \widehat{\mathrm{inj}}\big((F, v_1), \mathbf{n}; B_{k+1}(G_n, x_1)\big).$$

Given N and r, let $\mathcal{B}_r^{(N)}$ be the set of routed graphs on $\{1, 2, \dots, N\}$ that have radius r and contain exactly one of the representatives from each isomorphism class, and let $\mathcal{B}_r = \bigcup_{N=1}^{\infty} \mathcal{B}_r^{(N)}$. Then

$$\begin{aligned}\frac{1}{n}\mathrm{inj}(F, \mathbf{n}; G_n) &= \frac{1}{n}\sum_{x_1 \in V(G_n)} \widehat{\mathrm{inj}}\big((F, v_1), \mathbf{n}; B_{k+1}(G_n, x_1)\big) \\ &= \sum_{B \in \mathcal{B}_{k+1}} \widehat{\mathrm{inj}}\big((F, v_1), \mathbf{n}; B\big)\mathrm{Pr}_{x_1}\big(B_{k+1}(G_n, x_1) \sim B\big),\end{aligned}$$

where $\sim$ indicates rooted isomorphisms and the probability is the probability over rooted balls induced by the random choice of $x_1 \in V(G_n)$.

Since F is connected, $\widehat{\mathrm{inj}}((F, v_1), \mathbf{n}; B_{k+1}(G_n, x_1))$ is upper bounded by the constant $C = \max_{1\le i\le k}(n(i) + d_F(v_i))^{k-1}$. Therefore convergence in the sense of Benjamini–Schramm implies convergence of the right-hand side, giving that

$$\begin{aligned}\hat{t}(F,\mathbf{n}) &:= \lim_{n\to\infty}\frac{1}{|V(G_n)|}\,\mathrm{inj}(F,\mathbf{n};G_n)\\ &= \sum_{B\in\mathcal{B}_{k+1}}\widehat{\mathrm{inj}}\big((F,v_1),\mathbf{n};B\big)\Pr\big(B_{k+1}(G,x)\sim B\big)\\ &= E\big[\widehat{\mathrm{inj}}\big((F,v_1),\mathbf{n};(G,x)\big)\big],\end{aligned}\tag{39}$$

where $E[\cdot]$ denotes expectation over the random choices of the limit graph (G, x).

5.4.2. *Convergence in probability.* If G_n is a sequence of random graphs, the subgraph frequencies $\mathrm{inj}(F, \mathbf{n}; G_n,)$ are random numbers as well. Examining the last proof, one easily sees that the expectation of these numbers converges if G_n converges in the sense of Definition 2.1. For the preferential attachment graph, this gives

$$\lim_{n\to\infty}\frac{1}{|V(G_n)|}E\big[\mathrm{inj}(F,\mathbf{n};G_n)\big] = \hat{t}(F,\mathbf{n}),$$

where

$$\hat{t}(F,\mathbf{n}) = E\big[\widehat{\mathrm{inj}}\big((F,v_1),\mathbf{n};(T,0)\big)\big]\tag{40}$$

with $(T, 0)$ denoting the Pólya-point graph. It turns out that we can prove a little more, namely convergence in probability.

LEMMA 5.4. *Let G_n be one of the three versions of the preferential attachment graph defined in Section 2.1, let F be a finite connected graph and let $\mathbf{n} : V(F) \to \{0, 1, \dots, \}$. Then*

$$\frac{1}{n}\,\mathrm{inj}\big((F,\mathbf{n});G_n\big) \to \hat{t}(F,\mathbf{n}) \qquad \textit{in probability}.$$

PROOF. Assume that x_0 and x_0' are chosen independently uniformly at random from $V(G_n)$. Repeating the proof of Theorem 2.2, one easily obtains that the pair $((G_n, x_0), (G_n, x_0'))$ converges to two independent copies of the Pólya-point graph [more precisely, that the distribution of all pairs of balls $(B_r(G_n, x_0), B_r(G_n, x_0'))$ converges to the product distribution of the corresponding balls in $(T, 0)$]. As a consequence, the expectation of $[\frac{1}{n}\,\mathrm{inj}((F, \mathbf{n}); G_n)]^2$ converges $[\hat{t}(F, \mathbf{n})]^2$, which in turn implies the claim. $\square$

5.4.3. *Calculation of subgraph frequencies.* In this subsection, we calculate the limiting subgraph frequencies $\hat{t}(F, \mathbf{n})$ using the expression (40). Alternatively, one could use the intermediate expression in (39) and the fact that for each given rooted graph B of radius k, we can calculate the probability that the ball of radius k in the Pólya-point graph $(T, 0)$ is isomorphic to B. But this gives an expression involving the countably infinite sum over the balls in $\mathcal{B}_{k+1}$, while our calculation below only involves a finite number of terms.

In a preliminary step, we note that the Pólya-point graph $(T, 0)$ and the point process $\{x_{\bar{a}}\}$ can be easily recovered from the countable graph on $[0, 1]$ which is obtained by joining two points $x, x' \in [0, 1]$ by an edge whenever $x = x_{\bar{a}}$ and $x' = x_{\bar{a}'}$ for a pair of neighbors $\bar{a}, \bar{a}'$ in T. Identifying the point x_0 as the root, we obtain an infinite, random rooted tree on $[0, 1]$ which we will again denote by T.

Recalling (40), we will want to calculate the expected number of maps φ from $V(F)$ to $[0, 1]$ and are degree preserving homomorphism from $(F, \mathbf{n})$ into T that map v_1 into the root x_0. To this end, we explore the tree structure around the node x_0 in T, in a similar fashion as in Section 3.2. Obviously, if F is not a tree, then $\hat{t}(F, \mathbf{n}) = 0$. Otherwise, denote the vertex $v_1 \in V(F)$ as the root and obtain a rooted tree in which the set of children of every node is uniquely defined.

A mapping φ from vertices $v_1, v_2, \dots, v_k$ to points $x_1, x_2, \dots, x_k$ on the interval $[0, 1]$ defines a natural total order θ on $V(F)$. We say a mapping is consistent with total order θ if and only if for every i and j, $\theta(v_i) < \theta(v_j)$ implies $x_i < x_j$.

Given the positions $x_1, x_2, \dots, x_k$ (or equivalently the ordering θ), we can divide the children of every node v_i to two sets $L(v_i)$ and $R(v_i)$, depending on whether their corresponding points on the interval are to the left or right of x_i, respectively. With a slight abuse of notation, define $L = \bigcup_{1 \le i \le k} L(v_i)$ and $R = \bigcup_{1 \le i \le k} R(v_i)$. Note that $\{v_2, \dots, v_k\}$ is the disjoint union of L and R. Since we require that the degrees are preserved, the degree of a node x_i in T is $d_F(v_i) + n_i$. For the root $x_1 = x_0$ this gives $d_F(v_1) + n_1$ children, m to the left, and $n'_1 + |R(v_1)| = d_F(v_1) + n_1 - m$ to its right. If $v_i \in L$, its parent appears on its right. Therefore, of $n(v_i)$ remaining neighbors of x_i that are not mapped to any vertex in F, $n'(v_i) = d_F(v_i) + n(v_i) - (m + |R(v_i)| + 1)$ should appear to its right-hand side. For $v_i \in R$, $n'(v_i) = d_F(v_i) + n(v_i) - (m + |R(v_i)|)$.

Using the above notation, we can finally write the probability density function $p(F, \mathbf{n}, x)$ for a mapping from $V(F)$ to $x = (x_1, x_2, \dots, x_k)$ to be homomorphic and degree preserving. Conditioned on $\gamma(x_i) = \gamma_i$, it can be written as

$$
\begin{aligned}
&p(F, \mathbf{n}, x, \gamma) \\
&\quad = (\psi + 1) x_1^{\psi} \prod_{v_i \in V} \left(\frac{\exp(-H_i) H_i^{n'(i)}}{n'(i)!} \prod_{v_j \in L(v_i)} x_i^{-1} \prod_{v_j \in R(v_i)} \gamma_i \frac{\psi x_j^{\psi - 1}}{x_i^{\psi}} \right),
\end{aligned} \tag{41}
$$

where

$$
H_i = \gamma_i \frac{1 - x_i^{\psi}}{x_i^{\psi}}.
$$

The two inner product terms in the above equations are derived using the description of the Pólya-point in Section 2.3.2. The first term captures the probability that the remaining degree of x_i is the desired value $n'(i)$. Indeed, recalling that the children $x > x_i$ of a vertex x_i are given by a Poison process with density $\gamma_i \frac{\psi x^{\psi-1}}{x_i^{\psi}}$ on $[x_i, 1]$, we see that n'_i is a Poisson random variable with rate

$$\gamma_i \int_{x_i}^{1} \frac{\psi x^{\psi-1}}{x_i^{\psi}}\, dx = H_i,$$

giving the first term in the product above.

Also, γ_i is a Gamma variable with parameters $\alpha(i)$ and 1, where α_i depends on whether we discover v_i from right or left.

$$\alpha(i) = \begin{cases} m + 2mu + 1, & \text{if } v_i \in L, \\ m + 2mu, & \text{if } v_i \in R. \end{cases}$$

Similarly, $\alpha(1) = m + 2mu$. Let $C(\theta)$ be the simplex containing all points $x = (x_1, x_2, \dots, x_k)$ consistent with an ordering θ. Setting

$$\hat{t}(F, \mathbf{n}, \theta) = \int_{C(\theta)\times(0,\infty)^k} \prod_{i=1}^{k} \frac{e^{-\gamma_i}\gamma_k^{\alpha_k - 1}}{\Gamma(\alpha_i)} p(F, \mathbf{n}, x, \gamma)\, dx_1 \cdots dx_k\, d\gamma_1 \cdots d\gamma_k,$$

$t(F, \mathbf{n})$ can now be computed by summing $t(F, \mathbf{n}, \theta)$ over the $k!$ choices of θ.

Acknowledgment. We thank an anonymous referee for helping us in improving the presentation of the paper. The research was performed while N. Berger and A. Saberi were visiting Microsoft Research.

REFERENCES

[1] ALBERT, R. and BARABÁSI, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys.* **74** 47–97. MR1895096

[2] ALBERT, R., JEONG, H. and BARABÁSI, A. (1999). Diameter of the world wide web. *Nature* **401** 130–131.

[3] ALDOUS, D. (1998). Tree-valued Markov chains and Poisson–Galton–Watson distributions. In *Microsurveys in Discrete Probability* (*Princeton*, *NJ*, 1997). *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* **41** 1–20. Amer. Math. Soc., Providence, RI. MR1630406

[4] BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634

[5] BENJAMINI, I. and SCHRAMM, O. (2001). Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.* **6** 13 pp. (electronic). MR1873300

[6] BERGER, N., BORGS, C., CHAYES, J. T. and SABERI, A. (2005). On the spread of viruses on the internet. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 301–310. ACM, New York. MR2298278

[7] BOLLOBÁS, B. and RIORDAN, O. (2004). The diameter of a scale-free random graph. *Combinatorica* **24** 5–34. MR2057681

[8] BOLLOBÁS, B. and RIORDAN, O. (2011). Sparse graphs: Metrics and random models. *Random Structures Algorithms* **39** 1–38. MR2839983

[9] BOLLOBÁS, B. and RIORDAN, O. M. (2003). Mathematical results on scale-free random graphs. In *Handbook of Graphs and Networks* (S. Bornholdt and H. G. Schuster, eds.) 1–34. VCH, Weinheim. MR2016117

[10] BORGS, C., CHAYES, J., KAHN, J. and LOVÁSZ, L. (2013). Left and right convergence of graphs with bounded degree. *Random Structures Algorithms* **42** 1–28. MR2999210

[11] BORGS, C., CHAYES, J., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2006). Counting graph homomorphisms. In *Topics in Discrete Mathematics*. *Algorithms and Combinatorics* **26** 315–371. Springer, Berlin. MR2249277

[12] BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2008). Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** 1801–1851. MR2455626

[13] BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2012). Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math.* (2) **176** 151–219. MR2925382

[14] DURRETT, R. (1996). *Probability*: *Theory and Examples*, 2nd ed. Duxbury Press, Belmont, CA. MR1609153

[15] EGGENBERGER, F. and POLYA, G. (1923). Uber die statistik verketteter vorgange. *Zeitschrift für Angewandte Mathematik und Mechanik* **3** 279–289.

[16] ELEK, G. (2007). On limits of finite graphs. *Combinatorica* **27** 503–507. MR2359831

[17] FALOUTSOS, M., FALOUTSOS, P. and FALOUTSOS, C. (1999). On power-law relationships of the internet topology. In *SIGCOMM'99*: *Proceedings of the Conference on Applications*, *Technologies*, *Architectures*, *and Protocols for Computer Communication* 251–262. ACM, New York.

[18] LESKOVEC, J., CHAKRABARTI, D., KLEINBERG, J., FALOUTSOS, C. and GHAHRAMANI, Z. (2010). Kronecker graphs: An approach to modeling networks. *J. Mach. Learn. Res.* **11** 985–1042. MR2600637

[19] LESKOVEC, J., KLEINBERG, J. M. and FALOUTSOS, C. (2007). Graph evolution: Densification and shrinking diameters. *Transactions on Knowledge Discovery from Data* **1** Article 2.

[20] LOVÁSZ, L. and SZEGEDY, B. (2006). Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96** 933–957. MR2274085

[21] LYONS, R. (2005). Asymptotic enumeration of spanning trees. *Combin. Probab. Comput.* **14** 491–522. MR2160416

[22] RUDAS, A., TÓTH, B. and VALKÓ, B. (2007). Random trees and general branching processes. *Random Structures Algorithms* **31** 186–202. MR2343718

N. BERGER
MATHEMATICS DEPARTMENT
HEBREW UNIVERSITY
JERUSALEM 91904
ISRAEL
E-MAIL: berger@math.huji.ac.il

C. BORGS
J. T. CHAYES
MICROSOFT RESEARCH NEW ENGLAND
CAMBRIDGE, MASSACHUSETTS 02142
USA
E-MAIL: borgs@microsoft.com
chayes@microsoft.com

A. SABERI
MANAGEMENT SCIENCE AND ENGINEERING
STANFORD UNIVERSITY
PALO ALTO, CALIFORNIA 94305
USA
E-MAIL: saberi@stanford.edu