Check for updates

Adv. Appl. Probab. **57,** 1068–1100 (2025)
*doi:*10.1017/apr.2024.73

MULTIVARIATE REGULAR VARIATION OF PREFERENTIAL ATTACHMENT MODELS

ANJA JANSSEN,* ** AND
MAX ZIEGENBALG,* *** *Otto-von-Guericke University Magdeburg*

Abstract

We use the framework of multivariate regular variation to analyse the extremal behaviour of preferential attachment models. To this end, we follow a directed linear preferential attachment model for a random, heavy-tailed number of steps in time and treat the incoming edge count of all existing nodes as a random vector of random length. By combining martingale properties, moment bounds and a Breiman type theorem we show that the resulting quantity is multivariate regularly varying, both as a vector of fixed length formed by the edge counts of a finite number of oldest nodes, and also as a vector of random length viewed in sequence space. A Pólya urn representation allows us to explicitly describe the extremal dependence between the degrees with the help of Dirichlet distributions. As a by-product of our analysis we establish new results for almost sure convergence of the edge counts in sequence space as the number of nodes goes to infinity.

Keywords: Extreme value theory; multivariate regular variation; preferential attachment; random networks; random graphs

2010 Mathematics Subject Classification: Primary 60G70
Secondary 05C80, 60F25

1. Introduction

Preferential attachment random graphs, as initially introduced in [2, 7], have become prevalent models for the evolution of stochastic networks. The key feature of this class of models is that the current number of edges of a node affects the probability of a new node connecting to it. This underlying 'rich get richer'-dynamic is well-known to lead to a heavy-tailed behaviour in the resulting limit distribution of edge counts per node, as the number of nodes goes to infinity. To make this notion of heavy tails more precise, think of a simple preferential attachment model which consists at time $n = 0$ of a single node, labelled by 1, and a single directed edge from this node to itself, i.e. a loop. Now, at each following point $n = 1, 2, \ldots$ in time, we add a new node, labelled by $n + 1$, to the graph and attach one directed edge from this new node

Received 4 October 2023; accepted 10 December 2024.
* Postal address: Faculty of Mathematics, Institut für Mathematische Stochastik (IMST), Universitätsplatz 2, 39106 Magdeburg, Germany.
** Email address: anja.janssen@ovgu.de
*** Email address: max.ziegenbalg@ovgu.de

© The Author(s), 2025. Published by Cambridge University Press on behalf of Applied Probability Trust. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

to itself or another already existing node. This node to connect to is chosen at random, with a probability that depends on the number of incoming edges per node. More precisely, moving from time n to $n+1$ we create a new node that connects to the node with label $i \le n+2$ with the probability

$$p_i(n) := \frac{\deg_i^{\text{in}}(n) + \beta}{\sum_{k=1}^{n+2} (\deg_k^{\text{in}}(n) + \beta)}, \tag{1.1}$$

where $\deg_i^{\text{in}}(n)$, $i = 1, \dots, n+2$ stands for the in-degree of node i at time n, i.e. the number of incoming edges, and $\beta > 0$ denotes a so-called offset parameter. In this model, the resulting empirical in-degree distribution at time n, given by its probability mass function

$$i \mapsto p_n(i) := \frac{\sum_{k=1}^{n+1} \mathbb{1}_{\{\deg_k^{\text{in}}(n)=i\}}}{n+1}, \quad n \in \mathbb{N}, 1 \le i \le n+1, \tag{1.2}$$

can be shown to converge as $n \to \infty$ to a limiting probability mass function $i \mapsto p(i)$, $i \in \mathbb{N}$, such that there exists a constant $c > 0$ for which

$$\lim_{i \to \infty} \frac{p(i)}{i^{-(2+\beta)}} = c; \tag{1.3}$$

see [6]. We refer to Chapter 8 in [33] for a concise overview of preferential attachment models, covering both the motivation and a rigorous treatment of their asymptotic behaviour. For related work on limit theorems for degree counts, see [13, 20, 25, 31, 36]. In particular, we would like to point out Deijfen *et al.* [13] who introduce additional randomness in the model by considering a preferential attachment model, where, in each time step, a random heavy-tailed number of additional edges is drawn from the new vertex. With this paper we are following a similar idea, but instead we will evaluate the model after a random number of nodes has been added. More details of our model will be given below.

Equation (1.3) implies that the limiting probability mass function of the in-degrees decays like a power function. In network science, this property of a random graph is typically referred to as 'scale free' and it is often pointed out that a similar power-law behaviour is visible in the empirical degree distributions of real life networks; see [35] for a recent overview on the topic or see the KONECT database [22] for an abundance of real life examples.

One way to illustrate this phenomenon in a data set is by looking at the empirical quantile function of the in-degrees of an observed network. Figures 1 and 2 show so-called Zipf plots in the form of the ordered observations of in-degrees on a log–log scale (where rank 1 corresponds to the largest observation), for both a real data set and a simulated realisation of a preferential attachment model. The power-law behaviour manifests itself in a clear linear trend in the largest (and for the preferential attachment model, indeed all) observations.

However, this kind of asymptotic for the largest observations may not be sufficient to describe the relevant aspects of the extremal behaviour of a preferential attachment model, due to two shortcomings:

- Real networks are often evaluated after a random number of nodes have been added to them. For example, if a social network is evaluated at a given date, a random number of participants will have joined it by then. Now, the magnitude of an extremal event, such as the size of the largest degree, primarily depends on the number of participants, and so the distribution of largest degrees depends crucially on the random number of nodes observed.

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

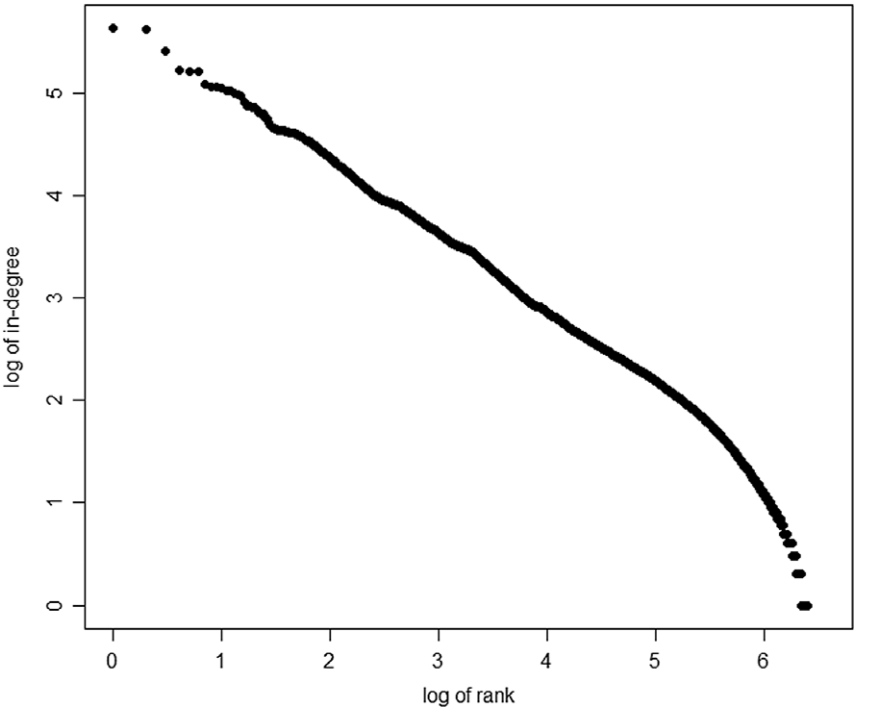


FIGURE 1. Zipf plot for in-degrees in the network of links between German Wikipedia articles; see [22], http://konect.cc/networks/wikipedia_link_de/.

- Neither the Zipf plot nor the asymptotic (1.3) provide any information about which nodes will eventually have the largest in-degree. Typically these are the oldest ones, but the exact order of the in-degrees is random and the individual magnitudes show an intricate dependence structure.

As a more concrete example, the asymptotic (1.3) is not sufficient to give an estimate for the probability that, say, the sum of the in-degrees of the four oldest nodes exceeds a certain high threshold and that these four in-degrees are in descending order.

We will come back to this question in our Example 2 below and will more generally provide an asymptotic analysis which allows us to address the two aforementioned shortcomings for a common class of preferential attachment models by using the framework of multivariate regular variation. To this end, we study a dynamic which is a generalisation of the two models analysed in [27], i.e. linear directed preferential attachment models in which a new node creates a fixed number of $l \in \mathbb{N}$ outgoing edges, and where we allow for an arbitrary offset parameter $\beta \geq 0$ as in (1.1). We then follow our network for a *random* number of steps, and assume that this number has a heavy-tailed distribution. It will turn out that both the index of regular variation of this heavy-tailed distribution and the parameters β and l of the network dynamic will influence the extremal behaviour of the in-degree vector.

As tools for our analysis, we extend approaches from [24, 27] to our more flexible class of models and combine them with a generalisation of a Breiman type result from [37] in order to derive the limiting quantities for the asymptotic behaviour of the network.

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

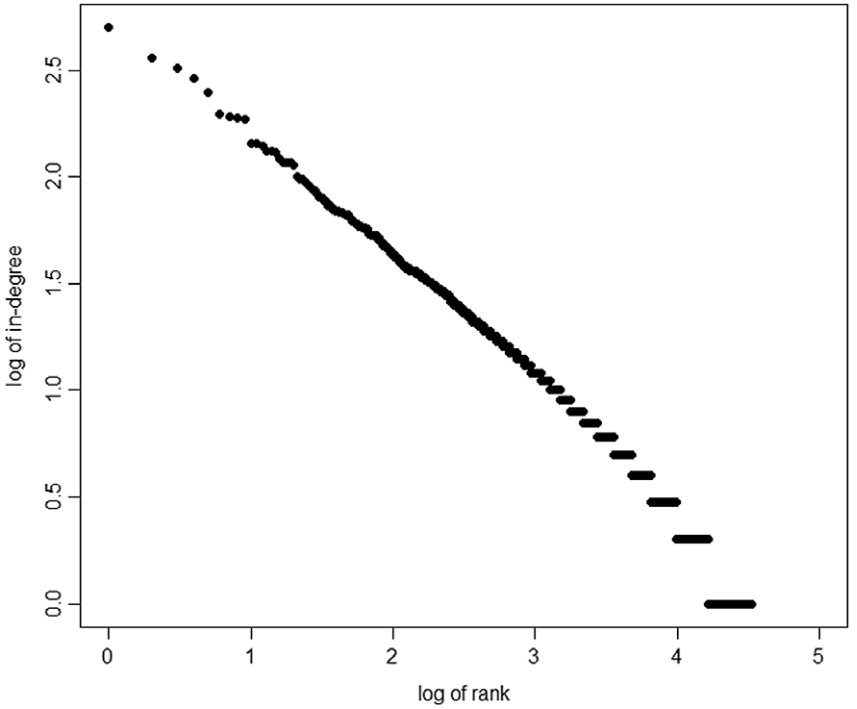


FIGURE 2. A Zipf plot for in-degrees of a simulated preferential attachment model after 100 000 time steps starting from one initial node with offset parameter $\beta = 1$. For nodes with large degrees (low rank) it shows strong similarities to the real life network from Figure 1.

The article is structured as follows. In Section 2 we introduce our model and make a connection to Pólya urns which allow us to apply martingale methods in order to derive an almost sure limit of mixed powers of finitely many in-degrees as the number of nodes goes to infinity (Proposition 1). Next, the concept of regular variation in Banach spaces is introduced in Section 3 along with a Breiman type result (Theorem 3) that allows us to transfer the regular variation of the observed number of nodes to regular variation of the in-degree vector. We first restrict the analysis to a finite stretch of the in-degree vector, i.e. we only analyse an arbitrary but fixed number of oldest nodes in the network and derive its regular variation (Theorem 4 and Corollary 1). In order to explicitly describe the extremal dependence structure we deduce the form of the so-called spectral measure. Again by making use of the Pólya urn representation and its limiting beta distributions, we can show that the spectral measure can be explicitly described with the help of Dirichlet distributions (Theorem 5 and Corollary 2). Subsequently, in Section 4, we extend our analysis to sequence space, which allows us for example to keep track of the extremal behaviour of the maximum degree of our network (Corollary 3). As an auxiliary result we first derive almost sure convergence, as the number of nodes tends to infinity, of the in-degree sequence in sequence space (Proposition 2). Finally, we arrive at a regular variation result in sequence space (Theorem 6). Longer proofs and auxiliary results have been deferred to Appendix A.

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

2. Background on preferential attachment models

2.1. Model definitions

In this section we introduce our random graph model and give an equivalent formulation of it that will turn out to be useful. Our model is a generalisation of the model considered in [27] in that we allow for an arbitrary offset parameter $\beta \geq 0$ in the linear preferential attachment function.

2.1.1. *The preferential attachment random graph* Our object of interest is a graph $G(n),\ n \in \mathbb{N}_0$, randomly evolving in a discrete time setting and growing by one vertex in each step from which a fixed amount $l \in \mathbb{N}$ of edges originate. As per usual, a graph is defined as the tuple $G(n) = (V(n), E(n))$ of its set of vertices and its set of edges. With $N(0)$ we denote the initial number of vertices and consequently with $N(n) = N(0) + n$ the number at time n. We will simply number these vertices in their order of appearance, so $V(n) = \{1, \ldots, N(0), N(0) + 1, \ldots, N(0) + n\}$, where the concrete numbering of the initial $N(0)$ vertices is of no interest. For the edges, we assume them to be directed and from now on we will only consider the in-degrees $\deg_k^{\text{in}}(n)$, $k \in N(n)$, of the vertex k at time n (as the out-degree is not random). If vertex k does not yet exist at time n, we set $\deg_k^{\text{in}}(n) = 0$. Finally, we want to introduce weights $D_k(n)$ which will determine the probability of a vertex k to get a new edge attached to it at time n. We deal with *linear* preferential attachment models and so let $f{:}\mathbb{N}_0 \to \mathbb{R}_{\geq 0}$ be an affine function with $f(x) = \alpha x + \beta,\ \alpha, \beta \geq 0$ and apply it to the (random) in-degrees: $D_k(n) := f(\deg_k^{\text{in}}(n))$. As the aforementioned probability shall be proportional to $D_k(n)$, without loss of generality we can restrict ourselves to weight functions f with $\alpha = 1$, i.e. set $D_k(n) = \deg_k^{\text{in}}(n) + \beta$, $k \in N(n)$. By choosing $\beta = l$ the weight $D_k(n),\ k > N(0)$ represents the total degree (in-degree + out-degree) of vertex k.

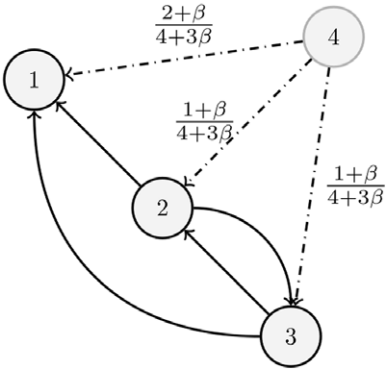
Next, we want to describe the random growth of the graph in more detail. At time 0 we start with an arbitrary but deterministic initial graph G(0) which is finite, meaning that $|V(0)| + |E(0)| < \infty$. When transitioning from time n to $n + 1$ a new vertex is added to the graph establishing a fixed number $l \in \mathbb{N}$ of edges to already existing vertices in two possible ways. The corresponding figures illustrate as an example how node 4 is added to an existing graph consisting of nodes 1, 2, 3 and edges (solid arrows) in each of the two models. Dashed-dotted arrows show all possibilities for creating a new edge with the corresponding probabilities. We consider two different models.

Model 0 (no loops): The first model does not allow for loops, i.e. an edge from a vertex to itself. We sequentially add the l edges, immediately updating the weights of the corresponding vertex. For an $i \in \{1, \ldots, l\}$ let $K_1, \ldots, K_{i-1}$ be the vertices chosen for the first $i - 1$ edges. The probability to attach the ith edge to vertex $k \in \{1, \ldots, N(n)\}$ is then

$$\frac{D_k(n) + \sum_{j=1}^{i-1} \mathbb{1}_{\{K_j = k\}}}{(i-1) + \sum_{i=1}^{N(n)} D_i(n)},$$

which is the proportion of the weight of k to the total sum of weights at this particular moment. Note that this formula does not consider the vertex $N(n) + 1$ we are about to add.

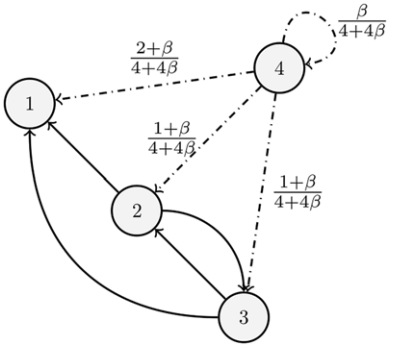
https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press



Model 1 (with loops): The second model allows for loops by a small modification of the above probability formula. To this end, we set the weight of the new vertex $D_{N(n)+1}(n) = \beta$. Then, with the same notation as in Model 0, we allow the vertex k to range over $\{1, \dots, N(n) + 1\}$ and use

$$\frac{D_k(n) + \sum_{j=1}^{i-1} \mathbb{1}_{\{K_j = k\}}}{(i-1) + \sum_{i=1}^{N(n)+1} D_i(n)}$$

as the probability for the ith edge to attach to it.



Thus, the difference between the two models is simply that in Model 1 the new vertex is added to the graph before its edges are attached to already existing ones, whereas in Model 0 it is added afterwards.

2.1.2. *An infinite-colour urn model* In order to analyse the limiting behaviour of preferential attachment models, it has proven to be beneficial to exploit a close relationship to generalised Pólya urns that allows the transfer of methods and results from the latter setting to the former. Both Pólya urns and preferential attachment models are random processes with reinforcement; see [28] for an overview on the topic. Starting with [3], Pólya urns have been fruitfully used to study the asymptotic behaviour of preferential attachment models; see also [4, 9, 15, 17, 27] and [34, Part II, Section 5] for related approaches. Our approach below is based on an almost sure representation of one process by the other, given in Lemma 1, which allows us to derive almost sure convergence statements, thereby complementing and extending results from [24, 27].

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

Urn model: The idea is that each colour k in an urn stands for a vertex of the random graph and the number of balls of that colour is the corresponding weight D_k. Since the weights can be any positive real number, we will also allow our urn to contain non-natural numbers of balls. So let us consider an urn that at time 0 contains an arbitrary (natural) number of balls of s different colours (representing the edges in the graph) plus β balls of each of those colours (the weight functions offset). Then, in the time step $n \mapsto n+1$, we randomly draw a ball from the urn, where the probability of a certain colour to be drawn, just like in a regular urn, is the ratio of the number of balls of that colour to the total number of balls. Afterwards we return the ball together with an additional one of the same colour. Furthermore, if n is a multiple of l, we add β more balls of the new colour $s + \frac{n}{l}$, a mechanic we will refer to as immigration. So one can view l time steps in the urn model as one step for the graph. We will denote the number of balls of colour k in the urn at time n by $C_k(n)$.

We obtain the following lemma describing the aforementioned connection.

Lemma 1. *Let $j \in \{0, 1\}$, $l \in \mathbb{N}$, $\beta > 0$ and in addition $N(0) \in \mathbb{N}$ and $(D_1(0), \ldots, D_{N(0)}(0)) \in \mathbb{R}_{\geq 0}^{N(0)}$. Then, there exists a probability space $(\Omega, \mathcal{A}, P)$ which accommodates two stochastic processes $(C_i(n))_{i,n\in\mathbb{N}}$ and $(D_i(n))_{i,n\in\mathbb{N}}$, such that*

- *$(D_i(n))_{i,n\in\mathbb{N}}$ has the same distribution as the random graph Model j described above (where $D_i(n)$ denotes the weight of vertex i after n new vertices have been added) with parameters l and β and a starting configuration given by N(0) vertices with weights $D_k(0)$, $k = 1, \ldots, N(0)$,*
- *$(C_i(n))_{i,n\in\mathbb{N}}$ has the same distribution as the urn model described above (where $C_i(n)$ denotes the number of balls of colour i after n draws have been performed) with parameters l and β and a starting configuration given by $s := N(0) + j$ colours, with $D_k(0)$ balls of colour $k = 1, \ldots, N(0)$ and, if $j = 1$, β balls of colour $N(0) + 1$,*

such that

$$(C_1(n \cdot l), \ldots, C_r(n \cdot l)) = (D_1(n), \ldots, D_r(n))$$

for any $n \in \mathbb{N}$, $r < s + n$.

Proof. Both the graph and the urn model for the first $s + n - 1$ colours/vertices observed after n steps in the respective model (where one step here means adding one ball of a certain colour/one edge to a certain node) form a Markov chain, and it can be easily checked that the transition probabilities for one step coincide. This ensures the existence of the underlying probability space as stated; see [19, Section 1.1].

Remark 1. The urn model provides several advantages for analysing the random graph in our context. First, it performs each step of adding an edge to the graph separately, i.e. in one draw from the urn, and second, it ignores additional structure in the graph which we are not interested in, such as the information related to which vertex is hanging on the other side of an edge. Another benefit of the urn model is the elimination of the need to distinguish between versions with and without loops when considering transition probabilities. We simply shift the starting amount of colours in the urn by j, which corresponds either to adding the vertex before attaching its edges or after. However, when we add the colour beforehand, it represents a vertex that should not yet exist, which is why we exclude the case $r = s + n$ in Lemma 1. In subsequent proofs, we will frequently alternate between the urn and graph models as permitted by this lemma.

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

A first observation we can make about the asymptotic behaviour of the individual weights/ball counts is that they will all tend to infinity almost surely.

Lemma 2. *Assume the graph/urn model of Lemma 1. Then, as $n \to \infty$, for arbitrary k the weight $D_k(n)$ of vertex k/ball count $C_k(n)$ of colour k diverges to infinity almost surely.*

Proof. We take the perspective of the urn model. Consider an arbitrary time $n = m \cdot l,\ m \in \mathbb{N}$, at which $C_k(n) > 0$, where such an n exists since we assumed $\beta > 0$. Let c denote the number of balls of colours other than k at this time. Then the probability that no further ball of colour k will be selected from here on is

$$\prod_{i=0}^{\infty}\Big(1 - \frac{C_k(n)}{c + i + \beta\lfloor \frac{i}{l} \rfloor}\Big).$$

The statement follows if we can show that this probability is equal to 0 or equivalently that the series

$$-\sum_{i=0}^{\infty} \ln\Big(1 - \frac{C_k(n)}{c + i + \beta\lfloor \frac{i}{l} \rfloor}\Big)$$

diverges to infinity. Since $\ln(1-x) \sim -x$ as $x \to 0$, the said series converges if and only if

$$\sum_{i=0}^{\infty} \frac{C_k(n)}{c + i + \beta\lfloor \frac{i}{l} \rfloor}$$

does so. As this expression can be bounded from below by the divergent harmonic series, the result follows.

2.2. The almost sure limit and its moments

By Lemma 2 all weights $D_k(n)$ diverge to infinity as time progresses, raising the question about the asymptotic behaviour of

$$D^r(n) := (D_1(n), \ldots, D_r(n)),$$

for an arbitrary but fixed length $r \in \mathbb{N}$ after suitable rescaling when $n \to \infty$. The following result shows the almost sure convergence and characterises the limiting vector in terms of its mixed moments.

Proposition 1. *Assume the graph/urn model of Lemma 1. Throughout this proposition the times n or m are assumed to be greater than or equal to $(r-s) \vee 0$ (in the random graph setting) or l times this number (in the urn setting). Define the generalised binomial coefficient*

$$\binom{x}{y} := \frac{\Gamma(x+1)}{\Gamma(y+1)\Gamma(x-y+1)} \quad \forall x, y \in \mathbb{R}$$

and let $k_1, \ldots, k_r \in \mathbb{R}_{\geq 0}$, $k := \sum_{i=1}^{r} k_i$.

(1) *There exists a normalising sequence $c(n,k) \sim n^{k \cdot l/(l+\beta)}$ such that, with respect to the natural filtration $\mathcal{F}_n := \sigma(C_k(m), m \leq n, k \leq s + \lfloor \frac{n}{l} \rfloor)$, the process*

$$\Big(\frac{1}{c(n,k)} \prod_{i=1}^{r} \binom{D_i(n) + k_i - 1}{k_i}, \mathcal{F}_{n \cdot l}\Big)_n \tag{2.1}$$

is a positive martingale.

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

(2) *In particular, the above process is almost surely convergent with limit*

$$\prod_{i=1}^{r} \frac{\zeta_i^{k_i}}{\Gamma(k_i+1)} \in L^1(\mathbb{P}), \tag{2.2}$$

where $\zeta_i = \lim_{n\to\infty} n^{-l/(l+\beta)} D_i(n)$.

(3) *The limit in (2) closes the martingale to the right, i.e.*

$$\mathbb{E}\Big(\prod_{i=1}^{r} \frac{\zeta_i^{k_i}}{\Gamma(k_i+1)} \Big| \mathcal{F}_{n\cdot l}\Big) = \frac{1}{c(n,k)} \prod_{i=1}^{r} \binom{D_i(n)+k_i-1}{k_i} \quad \text{for every } n \in \mathbb{N}.$$

Most notably this means (with $c(n,0)=1$*)*

$$\mathbb{E}\Big(\prod_{i=1}^{r} \zeta_i^{k_i}\Big) = \prod_{i=1}^{r} \Bigg[\Gamma(k_i+1) \cdot \frac{c((i-s)\vee 0, k_1+\ldots+k_{i-1})}{c((i-s)\vee 0, k_1+\ldots+k_i)} \cdot \begin{cases} \binom{D_i(0)+k_i-1}{k_i} & \text{if } i \le N(0) \\ \binom{\beta+k_i-1}{k_i} & \text{if } i > N(0) \end{cases} \Bigg] \tag{2.3}$$

The proof of this proposition is deferred to Appendix A.1.

3. Regular variation of the finite-dimensional weight vector

The starting point for the asymptotic analysis in Section 2.2 was to let the time index n go to infinity and derive properties of the almost sure limit of the weight vector after proper, time-dependent standardisation. In this section, we will take a different approach by focusing on the random vector

$$D^r(N) := (D_1(N), \ldots, D_r(N)), \tag{3.1}$$

of the weights of the oldest r nodes (including the ones already existing at time 0) evaluated at a *random* time N (where we set $D_k(N)=0$ if node $1 \le k \le r$ does not yet exist at time N) and how the extremes of this random vector can be described. We can think of N being the random number of steps that our network has already gone through at the time of observation. The extremal behaviour of $D^r(N)$ will thus both depend on the value of N and the network dynamics. In order to further analyse the behaviour, we first introduce a suitable framework.

3.1. Background on multivariate regular variation

When dealing with extremes of random vectors X we typically deem an outcome extreme if its norm $||X||$ exceeds a high threshold and we seek a stabilising behaviour after suitable normalisation if the threshold goes to infinity. This leads to the definition of multivariate regular variation, where the concept can be formalised in several slightly different ways. While the original definition was based on vague convergence, see [12, 30], we will employ the more flexible concept of $\mathbb{M}$-convergence, see [16, 21, 23], which will allow us to extend our results to the sequence space in Section 4.

Definition 1. *Let* $(B, ||\cdot||)$ *be a separable Banach space and set* $B^* := B \setminus \{0\}$. *Write* $\mathcal{B}(B^*)$ *for the Borel* σ*-algebra on* B^* *and let* $\mathbb{M}(B^*)$ *denote the space of measures on* $\mathcal{B}(B^*)$ *which are finite on sets bounded away from 0, i.e. sets* A *such that* $\inf_{a\in A} ||a|| > 0$. *Furthermore, write*

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

$\mathcal{C}(B^*)$ for the set of all non-negative, bounded and continuous functions f on B^* for which there exists an $\epsilon > 0$ such that f vanishes on $B_\epsilon(0) := \{b \in B : \|b\| < \epsilon\}$.

We say a random variable X with values in B is multivariate regularly varying in B^* if there exists a non-trivial measure $\mu \in \mathbb{M}(B^*)$, also called the limit measure, such that

$$\frac{\mathbb{P}(X/t \in \cdot)}{\mathbb{P}(||X|| > t)} \overset{t\to\infty}{\to} \mu(\,\cdot\,) \text{ in } \mathbb{M}(B^*), \tag{3.2}$$

where

$$\mu_t \to \mu \text{ in } \mathbb{M}(B^*) \quad \Leftrightarrow \quad \mu_t(f) := \int f \mathrm{d}\mu_t \to \int f \mathrm{d}\mu =: \mu(f) \quad \forall f \in \mathcal{C}(B^*). \tag{3.3}$$

The mode of convergence defined in (3.3) is called $\mathbb{M}$-convergence.

Remark 2. Definition 1 implies that, for any multivariate regularly varying X on $B \setminus \{0\}$, the random variable $||X||$ is (univariate) regularly varying on $\mathbb{R}_{>0}$ with some index α, meaning that

$$\lim_{t\to\infty} \frac{\mathbb{P}(||X|| > \lambda t)}{\mathbb{P}(||X|| > t)} = \lambda^{-\alpha} \quad \forall\, \lambda > 0.$$

For the remainder of this Section 3, we will focus on the special case $B = \mathbb{R}^n$ which also justifies the naming of *multivariate* regular variation. However, in preparation for Section 4, where we will consider the weight sequence as a whole, we will need to use the more general concept of a separable Banach space. This means that the choice of the norm matters in general (how much so, we will see in Section 4), while in $\mathbb{R}^n$, at least for qualitative considerations, it does not.

The limit measure μ has a certain structure implied by the multiplicative standardisation used in (3.2).

Theorem 1. (Theorem 3.1 in [16].) *Let X be a multivariate regularly varying random variable with limit measure μ. Then there exists an $\alpha > 0$ such that for all $\lambda > 0$ and $A \in \mathcal{B}(B^*)$ we have*

$$\mu(\lambda A) = \lambda^{-\alpha} \mu(A).$$

We call α the index of regular variation or tail index. To read α directly from the limit measure one may find the second of the following alternative characterisations of multivariate regular variation useful.

Theorem 2. (Corollary 4.4 in [23].) *Let X be a random variable in a separable Banach space B. Each of the following statements is equivalent to X being multivariate regularly varying with index α.*

(a) *There exist a non-trivial measure μ and an increasing function $b(t)$ such that*

$$t\mathbb{P}(X/b(t) \in \cdot) \overset{t\to\infty}{\to} \mu \text{ in } \mathbb{M}(B^*) \text{ and } \frac{b(\lambda t)}{b(t)} \overset{t\to\infty}{\to} \lambda^{\frac{1}{\alpha}} \ \forall \lambda > 0.$$

(b) *There exists a probability measure S on the unit sphere $\mathbb{S}_{||\cdot||} := \{x \in B : ||x|| = 1\}$, also called the spectral measure, such that*

$$\frac{\mathbb{P}(X/t \in \cdot)}{\mathbb{P}(||X|| > t)} \overset{t\to\infty}{\to} (\nu_\alpha \otimes S) \circ h^{-1} \text{ in } \mathbb{M}(B^*),$$

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

where ν_α is a measure on $\mathbb{R}_{>0}$ determined by its values on the right-unbounded intervals $\nu_\alpha(t,\infty) := t^{-\alpha}$ and $h:\mathbb{R}_{>0}\times\mathbb{S}_{||\cdot||}\to B$ with $h(x,y)=x\cdot y$ is the polar coordinate transformation.

Note that, despite the suggesting notation here, limit measures coming from different normalising functions $b(t)$ may differ up to a multiplicative factor. To obtain the limit measure from (3.2) in Definition 1 one may choose the normalising function $b(t) := F^{\leftarrow}_{||X||}(1-t^{-1})$, where $F^{\leftarrow}_{||X||}(u) := \inf\{x\in\mathbb{R}:\mathbb{P}(||X||\le x)\ge u\}$ stands for the generalised inverse of the distribution function of $||X||$.

3.2. Main result

In this section we will start with our analysis of the asymptotic behaviour of the weight vector in the framework of multivariate regular variation. To this end, consider first for fixed n, $r\in\mathbb{N}$ the random vector $D^r(n)=(D_1(n),\dots,D_r(n))$, where we set $D_k(n)=0$ if $k>N(0)+n$. One immediately finds a deterministic upper bound for the ℓ_1-norm of the vector in

$$||D^r(n)||_1 \le ||D^r(0)||_1 + n\cdot(l+\beta),$$

which in view of Remark 2 rules out multivariate regular variation of $D^r(n)$. We thus need the number of nodes in the system to be random in order to witness heavy-tailed behaviour of the weight vector. Viewing the number of nodes or, equivalently, the number of completed steps in the evolution of a network as random is also reasonable from a modelling perspective. We typically do not observe a network after an a priori known number of steps, especially if new nodes are added according to some random arrival process instead of regularly over time. The results in Section 2.2 imply that the number of nodes is the driving factor behind an extremal behaviour of node weights. Our approach is thus to assume a heavy-tailed behaviour of N, which we choose to be (univariate) regularly varying, and interpret very large real life networks as the extremal realisations of N. Many typical regularly varying distributions are continuous, such as the Pareto or Student-t distributions. However, if we round any regularly varying random variable Y to $\lfloor Y\rfloor$, the result remains regularly varying, leading to a wide variety of possible distributions for $N\in\mathbb{N}_0$.

Now, an often observed principle in extreme value theory is derived from Breiman's lemma and applies in settings in which, roughly speaking, a system consists of a heavy-tailed component and a light-tailed one. Then, the heavy-tailed component will typically drive the extremal behaviour of the system and determine its index of regular variation, while the light-tailed one may well have an impact on the extremal dependence, in terms of the form of the spectral measure. It turns out that in our case a similar Breiman result, more precisely an extension of Theorem 3 in [37] adapted to Banach space-valued processes, applies as well.

Theorem 3. *Let $(X_t)_{t\in\mathbb{R}_{\ge0}}$ be an at least one-sidedly continuous stochastic process with values in a separable Banach space $(B,||\cdot||)$ and let T be a positive random variable, such that the following conditions are satisfied.*

(1) *T is regularly varying with index $\alpha>0$, i.e. for some scaling function $b(t)$ we have*

$$t\cdot\mathbb{P}\Big(\frac{T}{b(t)}>\lambda\Big)\overset{t\to\infty}{\to}\lambda^{-\alpha}=\nu_\alpha((\lambda,\infty))\quad\forall\lambda>0.$$

(2) *(X_t) and T are independent, i.e. $\sigma(T)\perp\sigma(X_t,t\ge0)$.*

(3) *X_t converges to some $X_\infty\in B^*$ almost surely as $t\to\infty$.*

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

(4) *For some $\alpha' > \alpha$, the following moment condition holds:*

$$\sup_{t\geq 0} \mathbb{E}\big(||X_t||^{\alpha'}\big) < \infty.$$

Then $T \cdot X_T$ is multivariate regularly varying in B^* with index α and

$$t\mathbb{P}\Big(\frac{T \cdot X_T}{b(t)} \in \cdot\Big) \overset{t\to\infty}{\to} [\nu_\alpha \otimes \mathbb{P}(X_\infty \in \cdot)] \circ \tilde{h}^{-1} \text{ in } \mathbb{M}(B^*), \tag{3.4}$$

where $\tilde{h}$:$\mathbb{R}_{>0} \times B \to B$, $\tilde{h}(x, y) = x \cdot y$.

The proof of this theorem is deferred to Appendix A.2.

Using this we can now state and prove our findings.

Theorem 4. *Consider a preferential attachment random graph with the notation as introduced in the last sections. In particular denote with l and β the number of edges added with each new vertex and the weight functions offset, respectively. For fixed n, $r \in \mathbb{N}$, let $D^r(n)$:= $(D_1(n), \ldots, D_r(n))$ be the corresponding weight vector and let N be a positive integer-valued random variable. If the assumptions*

(1) *N and $(D^r(n))_{n\in\mathbb{N}}$ are independent and*

(2) *N is regularly varying with index $\alpha > 0$*

are satisfied, then $D^r(N)$ is multivariate regularly varying with index $\alpha \cdot \frac{l+\beta}{l}$ for all $r \in \mathbb{N}$.

Proof. In order to apply Theorem 3 consider the process (X_t),

$$X_t := 0,\ t < 1, \qquad X_t := \frac{1}{t} \cdot D^r\big(\big\lceil t^{\frac{l+\beta}{l}} \big\rceil\big),\ t \geq 1,$$

and the random variable T with

$$T := N^{\frac{l}{l+\beta}}.$$

Then $T \cdot X_T = D^r(N)$. Therefore, all we need to do is to check the conditions from Theorem 3 and determine the index of regular variation.

Conditions (1) and (2): The independence follows immediately from assumption 1. As for the regular variation we consider the equivalent formulation of assumption 2,

$$\frac{\mathbb{P}(N > x\lambda)}{\mathbb{P}(N > x)} \to \lambda^{-\alpha}, \quad x \to \infty,$$

which implies

$$\frac{\mathbb{P}(T > x\lambda)}{\mathbb{P}(T > x)} = \frac{\mathbb{P}(N > x^{(l+\beta)/l}\lambda^{(l+\beta)/l})}{\mathbb{P}(N > x^{(l+\beta)/l})} \to \lambda^{-\alpha\cdot\frac{l+\beta}{l}}, \quad x \to \infty,$$

i.e. T is regularly varying with index $\alpha \cdot \frac{l+\beta}{l}$.

Condition (3): After the transformation $t' := t^{l/(l+\beta)}$ we apply Lemma 1 to

$$X_{t'} = \frac{1}{t^{l/(l+\beta)}} D^r(\lceil t \rceil) = \Big(\frac{\lceil t \rceil}{t}\Big)^{\frac{l}{l+\beta}} \frac{1}{\lceil t \rceil^{l/(l+\beta)}} D^r(\lceil t \rceil), \quad t \geq 1,$$

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

and the almost sure convergence follows.

Condition (4): Because of the definition of (X_t) we will only consider $t \geq 1$ and we set again $t' := t^{\frac{l}{l+\beta}}$. For $q \in \mathbb{N}$ we have by Lemma 1 and the moment bounds from Lemma 7 from the Appendix that

$$\begin{aligned}\mathbb{E}(||X_{t'}||_1^q) &= \frac{1}{t^{ql/(l+\beta)}}\mathbb{E}([D_1(\lceil t\rceil) + \ldots + D_r(\lceil t\rceil)]^q) \\ &= \frac{1}{t^{ql/(l+\beta)}}\mathbb{E}([C_1(l\cdot\lceil t\rceil) + \ldots + C_r(l\cdot\lceil t\rceil)]^q) \\ &\leq C\cdot\frac{\lceil t\rceil^{ql/(l+\beta)}}{t^{ql/(l+\beta)}} \leq C\cdot 2^{\frac{ql}{l+\beta}}\end{aligned}$$

for some $C > 0$, which completes the verification of the conditions of Theorem 3.

For the actual in-degree vector without the weight functions offset, we obtain the following corollary.

Corollary 1. *Under the same assumptions as in Theorem 4, the vector $D^r(N) - \beta$ is multivariate regularly varying with the same index of regular variation and the same limit measure as $D^r(N)$.*

Proof. This is a consequence of Lemma 3.12 from [18].

3.3. Characterisation of the spectral measure

Having verified the multivariate regular variation of the vector $D^r(N)$, the natural next step is to characterise its spectral measure S. Over the next few lemmas we shall once more switch into the urn perspective for this purpose. Because of the close relation to Pólya urns the following beta and beta-related distributions will come up, which we repeat here for convenience and since there exist different parameterisations in the literature.

Definition 2

(1) A random variable on $[0, 1]$ is said to follow a beta distribution $\mathsf{Beta}(a, b)$, with parameters $a, b > 0$, if it has a density function

$$f(x) = \frac{1}{\mathrm{B}(a,b)}x^{a-1}(1-x)^{b-1}\cdot\mathbb{1}_{[0,1]}(x),$$

where $\mathrm{B}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ is the beta function.

(2) A random vector with values in the standard $(r-1)$-simplex in $\mathbb{R}^r_{\geq 0}$, i.e. the set $\Sigma^{r-1} := \{x = (x_1, \ldots, x_r) \in \mathbb{R}^r_{\geq 0} : x_1 + \ldots + x_r = 1\}$, is said to follow a Dirichlet distribution $\mathsf{Dir}(a_1, \ldots, a_r)$ with parameters $a_1, \ldots, a_r > 0$ if it has a density function

$$f(x) = \frac{1}{\mathrm{B}(a_1,\ldots,a_r)}\prod_{i=1}^r x_i^{a_i-1}\cdot\mathbb{1}_{\Sigma^{r-1}}(x),$$

where $\mathrm{B}(a_1, \ldots, a_r) := \frac{\Gamma(a_1)\cdots\Gamma(a_r)}{\Gamma(a_1+\ldots+a_r)}$ is the multivariate beta function.

(3) A random vector with values in the standard $(r-1)$-simplex in $\mathbb{R}^r_{\geq 0}$ is said to follow a generalised Dirichlet distribution $\mathsf{GDir}(a_1, b_1, \ldots, a_{r-1}, b_{r-1})$ with parameters

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

$a_1, b_1, \ldots, a_{r-1}, b_{r-1} > 0$ if it has a density function

$$f(x) = \frac{1}{\prod_{i=1}^{r-1} \mathrm{B}(a_i, b_i)} x_r^{b_{r-1}-1} \prod_{i=1}^{r-1} \Big(x_i^{a_i-1} \Big(\sum_{j=i}^{r} x_j \Big)^{b_{i-1}-(a_i+b_i)} \Big) \cdot \mathbb{1}_{\Sigma^{r-1}}(x),$$

where b_0 is arbitrary.

Remark 3. The generalised Dirichlet distribution boils down to the standard version if its parameters satisfy $b_{i-1} = a_i + b_i$. Its density function then depends only on the parameters $a_1, \ldots, a_{r-1}$ and b_{r-1} which can be viewed as the parameters $a_1, \ldots, a_r$ of the standard Dirichlet distribution. These two distributions are also connected to the beta distribution through a stick breaking experiment, in which a random vector is constructed by breaking independent, beta-distributed fractions of a unit length stick for a fixed number of times, and the lengths of the resulting pieces form the components of the random vector. In general, this approach leads to a generalised Dirichlet distribution with parameters equal to those of the beta-distributed fractions in the order in which they were broken off (see [10]). To obtain a standard Dirichlet vector, the beta distributions must satisfy the parameter constraints mentioned above.

We begin with a well-known result for traditional multi-colour Pólya urns, i.e. our urn model but without immigration. The proof can be found in [5].

Lemma 3. *Consider an urn with balls of $r \in \mathbb{N}$ different colours and starting amounts $a^r = (a_1, \ldots, a_r) \in \mathbb{R}^r_{>0}$. Let $(X(n))_{n\in\mathbb{N}}$ be the sequence of colours drawn in each time step by following a traditional multi-colour Pólya urn scheme, where we add one ball of the drawn colour in each step. Then, for each colour i, the proportions of ball counts*

$$\frac{a_i + \sum_{k=1}^{n} \mathbb{1}_{\{i\}}(X(k))}{\sum_{j=1}^{r} a_j + n}, \quad i = 1, \ldots, r$$

converge, for $n \to \infty$, almost surely to some random variable Y_i such that

$$Y^r := (Y_1, \ldots, Y_r) \overset{\mathrm{d}}{=} \mathsf{Dir}(a_1, \ldots, a_r).$$

Notation 1. *In order to describe the spectral measure it is convenient to look at the projection of the limiting vector from Proposition 1(2), i.e.*

$$\zeta^r := (\zeta_1, \ldots, \zeta_r) = (\lim_{n\to\infty} n^{-l/(l+\beta)} D_i(n))_{1\le i\le r},$$

on the ℓ_1-unit sphere along with some derived proportions. We are specifically interested in two orderings of this vector: forwards,

$$S^r := (S_1, \ldots, S_r) := (\zeta_1, \ldots, \zeta_r)/(\zeta_1 + \ldots + \zeta_r), \text{ with}$$
$$B_k := \frac{S_k}{S_k + \ldots + S_r} = \frac{\zeta_k}{\zeta_k + \ldots + \zeta_r}, \quad 1 \le k \le r,$$

and backwards,

$$S^r_{\downarrow} = (S_r, \ldots, S_1), \text{ with } B^{\downarrow}_k := \frac{S_k}{S_1 + \ldots + S_k} = \frac{\zeta_k}{\zeta_1 + \ldots + \zeta_k}, \quad 1 \le k \le r.$$

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

In order to apply Lemma 3 to our problem we need to observe that once all colours we want to consider are present in the urn (which is at time $n_0^r = l(r-s) \vee 0$), these colours, considered separately from the rest of the urn, will behave just like they would in a traditional multi-colour Pólya urn. We thus obtain the following mixture distribution for S^r.

Theorem 5. *Let $r \in \mathbb{N}$, $C^r(n) := (C_1(n), \ldots, C_r(n))$, $n \in \mathbb{N}$, be the vector of the first r colours in the urn model after n steps with s initial colours from Lemma 1 and let S^r be, as in Notation 1, the ℓ_1-projection of the limiting vector in the corresponding random graph model.*

(1) *The limiting vector S^r follows a mixture of Dirichlet distributions, i.e.*

$$S^r \stackrel{\mathrm{d}}{=} \sum_{c \in \mathbb{R}^r_{\geq 0}} \mathbb{P}(C^r(n_0^r) = c) \cdot \mathsf{Dir}(c_1, \ldots, c_r).$$

(2) *For the vector obtained by inverting the order of components in S^r we additionally have*

$$S^r_{\downarrow} \stackrel{\mathrm{d}}{=} \mathsf{GDir}(a_r, b_r, \ldots, a_2, b_2),$$

where

$$(a_k, b_k) = \begin{cases} (C_k(0), \sum_{i=1}^{k-1} C_i(0)) & \text{if } 2 \leq k \leq s, \\ (\beta, \sum_{i=1}^{s} C_i(0) + (l+\beta)(k-1-s) + l) & \text{if } k > s. \end{cases}$$

The proof relies on the following lemma.

Lemma 4. *Let B_k, $B_k^{\downarrow}$, $1 \leq k \leq r$, be as in Notation 1. We have:*

(1) $B_1^{\downarrow} \equiv 1$, $B_k^{\downarrow} \stackrel{\mathrm{d}}{=} \begin{cases} \mathsf{Beta}(C_k(0), \sum_{i=1}^{k-1} C_i(0)) & \text{if } 2 \leq k \leq s, \\ \mathsf{Beta}(\beta, \sum_{i=1}^{s} C_i(0) + (l+\beta)(k-1-s) + l) & \text{if } k > s; \end{cases}$

(2) *the $B_1^{\downarrow}, \ldots, B_r^{\downarrow}$ are independent;*

(3) *both $(B_1, \ldots, B_r)$ and $(B_1^{\downarrow}, \ldots, B_r^{\downarrow})$ are independent of $\zeta_1 + \ldots + \zeta_r$.*

The proof of this lemma is deferred to Appendix A.4.

Proof of Theorem 5.

(1) We start by conditioning the urn model on its composition at time n_0^r, which is how the mixture distributions in the statement arise. Given that $C^r(n_0^r) = (c_1, \ldots, c_r) \in \mathbb{R}^r_{\geq 0}$, let t_i be the random time where for the ith time a ball from one of the colours 1 to r is chosen. At these times the vector of ratios of balls of each colour to the total number of balls with colours 1 to r behaves just like in a traditional multi-colour Pólya urn and therefore, by Lemma 3, converges almost surely to a Dirichlet distribution with $c_1, \ldots, c_r$ as parameters. At the same time, however, it converges (conditionally on $C^r(n_0^r) = (c_1, \ldots, c_r)$) to S^r, and so the statement follows.

(2) We write

$$S^r_{\downarrow} = \left(B_r^{\downarrow}, (1 - B_r^{\downarrow}) B_{r-1}^{\downarrow}, \ldots, \left(\prod_{i=2}^{r} (1 - B_i^{\downarrow}) \right) B_1^{\downarrow} \right),$$

and since, by Lemma 4, the $B_i^{\downarrow}$ are independent and have beta distributions, we can apply Remark 3 to arrive at the generalised Dirichlet distribution with the given parameters.

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

With these tools, we are now equipped to characterise the spectral measure.

Corollary 2. *The spectral measure of regular variation for the random vectors* $(D_1(N), \ldots, D_r(N))$ *and* $(D_r(N), \ldots, D_1(N))$*, with respect to the norm* $||\cdot||_1$*, coincides with the distribution of* S^r *and* $S^r_{\downarrow}$*, respectively, from Theorem 5.*

Proof. We restrict ourselves to the forwards case, i.e. focus on $(D_1(N), \ldots, D_r(N)) = D^r(N)$. By (3.4) for the limit measure μ provided in Theorem 3 and with X_∞ given by ζ^r we get

$$\mu = (\nu_\alpha \otimes \mathbb{P}(\zeta^r \in \cdot)) \circ \tilde{h}^{-1},$$

with $\tilde{h}:\mathbb{R}_{>0} \times \mathbb{R}^r_{\geq 0} \to \mathbb{R}^r_{\geq 0}$, $\tilde{h}(x, y) = x \cdot y$ and $\zeta^r = (\zeta_1, \ldots, \zeta_r)$. On the other hand, by Theorem 2(b) the corresponding spectral measure $S(A)$, for an S-continuity Borel set $A \subset \mathbb{S}_{\|\cdot\|}$, is given by

$$\begin{aligned} S(A) &= \lim_{t\to\infty} \frac{\mathbb{P}\left(\frac{D^r(N)}{\|(D^r(N)\|_1} \in A,\ \|(D^r(N)\|_1 > t\right)}{\mathbb{P}(\|(D^r(N)\|_1 > t)} \\ &= \frac{\mu(\{x \in \mathbb{R}^r_{\geq 0}: \|x\|_1 > 1,\ x/\|x\|_1 \in A\})}{\mu(\{x \in \mathbb{R}^r_{\geq 0}: \|x\|_1 > 1\})}. \end{aligned}$$

Now,

$$\begin{aligned} &\mu(\{x \in \mathbb{R}^r: \|x\|_1 > 1,\ x/\|x\|_1 \in A\} \\ &= \int_{\mathbb{R}^r_{\geq 0}} \int_{\mathbb{R}_{>0}} \mathbb{1}_{\{x\cdot\|y\|_1 > 1\}} \mathbb{1}_{\{\frac{y}{\|y\|_1} \in A\}}\, \nu_\alpha(\mathrm{d}x)\, \mathbb{P}(\zeta^r \in \mathrm{d}y) \\ &= \int_{\mathbb{R}^r_{\geq 0}} \|y\|_1^\alpha \mathbb{1}_{\{\frac{y}{\|y\|_1} \in A\}}\, \mathbb{P}(\zeta^r \in \mathrm{d}y) \\ &= \mathbb{E}(\|\zeta^r\|_1^\alpha) \cdot \mathbb{P}\Big(\frac{\zeta^r}{\|\zeta^r\|_1} \in A\Big), \end{aligned}$$

where the last equality follows from the independence of $\|\zeta^r\|_1 = \zeta_1 + \ldots + \zeta_r$ and $\frac{\zeta^r}{\|\zeta^r\|_1} = f(B_1, \ldots, B_r)$ (statement (3) in Lemma 4), with $f = (f_1, \ldots, f_r)$ given by

$$f_i(B_1, \ldots, B_r) = B_i \prod_{j=1}^{i-1} (1 - B_j), \quad i = 1, \ldots, r.$$

Going back to determining S, we thus get

$$S(A) = \frac{\mathbb{E}(\|\zeta^r\|_1^\alpha) \cdot \mathbb{P}(\frac{\zeta^r}{\|\zeta^r\|_1} \in A)}{\mathbb{E}(\|\zeta^r\|_1^\alpha)} = \mathbb{P}\Big(\frac{(\zeta_1, \ldots, \zeta_r)}{\zeta_1 + \ldots + \zeta_r} \in A\Big) = \mathbb{P}(S^r \in A),$$

and the statement follows.

We conclude this section with an example that demonstrates how the previous results can be applied to approximate the probabilities of extremal events, starting with a general approach and followed by a specific numerical example.

Example 1. We can use the limit/spectral measure to approximate conditional probabilities given large exceedances of the weight vector. A simple special case is to condition on

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

$\{\|D^r(N)\|_1 > t\}$ for some large t. Let A be a set in $\mathcal{B}(\mathbb{R}^r_{\geq 0} \setminus \{0\})$ bounded away from 0 and with $\mu(\partial A) = 0$. Then we have, using Theorem 2(b) (and the polar coordinate function h as defined there),

$$\begin{aligned}\mathbb{P}(D^r(N) \in tA \mid \|D^r(N)\|_1 > t) &= \frac{\mathbb{P}(D^r(N)/t \in A \cap \{x{:}\|x\|_1 > 1\})}{\mathbb{P}(\|D^r(N)\|_1 > t)} \\ &\approx (\nu_\alpha \otimes S) \circ h^{-1}(A \cap \{x{:}\|x\|_1 > 1\}),\end{aligned}$$

which does not depend on t anymore. So, as long as we know that our r vertices of interest have received a large amount of edges, we can approximate above probability using the limit/spectral measure.

Next, let us assume that A only contains information about the proportions of the $D_1(N), \ldots, D_r(N)$, i.e. that A is of the form $\{r \cdot \theta \mid (r, \theta) \in (1, \infty) \times A^*\}$ with $A^* \in \mathcal{B}(\mathbb{S}_{\|\cdot\|_1})$; then

$$\begin{aligned}\mathbb{P}(D^r(N) \in tA \mid \|D^r(N)\|_1 \geq t) &\approx (\nu_\alpha \otimes S) \circ h^{-1}(A \cap \{x{:}\|x\|_1 > 1\}) \\ &= (\nu_\alpha \otimes S) \circ h^{-1}(\{r \cdot \theta{:} r > 1, \theta \in A^*\}) \\ &= S(A^*).\end{aligned}$$

So, given that the total weight of our vector $D^r(N)$ is large, the proportions of its components follow roughly a mixture of Dirichlet distributions or a reversed generalised Dirichlet distribution.

Example 2. Following up on Example 1, let us consider a concrete extremal event. We want to determine the probability that the first four vertices are extreme with respect to the $||\cdot||_1$-norm, while also being in descending order, i.e. with

$$A = \{r \cdot \theta \mid (r, \theta) \in (1, \infty) \times A^*\} \text{ for } A^* := \{(x_1, x_2, x_3, x_4) \in \mathbb{S}_{\|\cdot\|_1} \mid x_1 \geq x_2 \geq x_3 \geq x_4\}$$

we want to approximate

$$\mathbb{P}(D^4(N) \in tA, ||D^4(N)||_1 > t) = \mathbb{P}(||D^4(N)||_1 > t) \cdot \mathbb{P}(D^4(N) \in tA \mid ||D^4(N)||_1 > t).$$

The second factor can be approximated by the spectral measure as explained in Example 1. For the first one, we get, using Theorem 3,

$$\frac{\mathbb{P}(||D^4(N)||_1 > t)}{\mathbb{P}(N^{l/(l+\beta)} > t)} = u(t) \cdot \mathbb{P}\Big(\frac{||D^4(N)||_1}{b(u(t))} > \frac{t}{b(u(t))}\Big) \overset{t\to\infty}{\to} \mathbb{E}\big(||\zeta^4||_1^{\alpha \cdot \frac{l+\beta}{l}}\big),$$

where $u(t) := \mathbb{P}(N^{\frac{l}{l+\beta}} > t)^{-1}$ and $b := u^{\leftarrow}$ is the generalised inverse of the monotone function u with $b(u(t)) \sim t$; see [11] B1.9 10. Overall, we obtain for large t,

$$\mathbb{P}(D^4(N) \in A, ||D^4(N)||_1 > t) \approx \mathbb{P}(N^{l/(l+\beta)} > t) \cdot \mathbb{E}\big(||\zeta^4||_1^{\alpha \cdot \frac{l+\beta}{l}}\big) \cdot S(A^*). \tag{3.5}$$

To get an impression of the quality of the approximation, we present the results of a small simulation study. The random value of N in our simulations is given by $\lfloor Y \rfloor$, where Y follows a Pareto(α) distribution with $\alpha = 1$, i.e. $P(N \geq k) = k^{-1}$, $k \in \mathbb{N}$. Table 1 shows values of the above approximation (3.5) for different choices of model parameters l and β and threshold t,

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

TABLE 1. Comparison of the left- and right-hand sides of (3.5) for several parameter constellations. The left-hand side was approximated by empirical probabilities based on 10^7 realisations of the network.

l	β	t	Empirical probability	Approximation
1	1	150	0.00968%	0.00949%
3	1	500	0.06325%	0.06215%
3	3	500	0.01272%	0.01255%

where $\mathbb{E}\big(||\xi^4||_1^{\alpha\cdot\frac{l+\beta}{l}}\big)$ follows from (2.3) and $S(A^*)$ has been evaluated by numerical integration with respect to the density of the generalised Dirichlet distribution. Since no closed expressions exist for the true probabilities, we replace them by empirical probabilities derived from 10^7 realisations of the network.

4. Generalisation to sequence spaces

4.1. Motivation and framework

Example 2 has shown that the results developed in the previous section allow us to approximate probabilities of extreme events—but with the restriction that those extreme events may only depend on a fixed number r of the first nodes in the graph. For several natural applications this approach is then not sufficient, as we would for example like to approximate probabilities of events involving the maximum degree of all existing nodes. In this section, we will extend our scope to include the asymptotic behaviour of the entire weight process rather than just its first r vertices. We will thus again work in the framework of regular variation as in Section 3.1, but with a suitably adjusted Banach space B for sequences. The regular variation of random sequences has previously been studied in [32]. In order to find an appropriate B we start with the space of all finite sequences:

$$c_{00} := \{x = (x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \mid \exists N \in \mathbb{N}\ \forall k \geq N : x_k = 0\}.$$

At any time n our weight process as well as the urn process can be represented as elements of this space:

$$D(n) := (D_1(n), \dots, D_{N(0)}(n), D_{N(1)}(n), \dots, D_{N(n)}(n), 0, \dots),$$
$$C(n) := (C_1(n\cdot l), \dots, C_s(n\cdot l), C_{s+1}(n\cdot l), \dots, C_{s+n}(n\cdot l), 0, \dots).$$

A crucial assumption in Theorem 3 is the almost sure convergence of the process and so we set B equal to the completion of c_{00} with respect to a norm $\|\cdot\|$ on $\mathbb{R}^{\mathbb{N}}$, i.e.

$$B = c_{\|\cdot\|} := \big\{x \in \mathbb{R}^{\mathbb{N}} \mid \|x\| < \infty \text{ and } \exists\, x_n \in c_{00},\, n \in \mathbb{N} : \lim_{n\to\infty} \|x - x_n\| = 0\big\}.$$

In contrast to the previously studied finite-dimensional setting, different norms on sequence spaces are no longer equivalent, which means that in order to study regular variation we have to find a suitable norm $\|\cdot\|$ for our model, where we restrict ourselves to the ℓ_p-norms. Then, by construction, B is a separable Banach space which allows us to employ the framework from Section 3.1 again.

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

4.2. The Breiman conditions in sequence space

Our goal is to establish a result similar to Theorem 4 again by virtue of Breiman's lemma. To this end, we need to check that conditions (3) and (4) of Theorem 3 are satisfied, which is shown in the following proposition.

Proposition 2. *Assume the graph/urn model of Lemma 1 and let*

$$\zeta := (\zeta_i)_{i\in\mathbb{N}} = \left(\lim_{n\to\infty} \frac{D_i(n)}{n^{l/(l+\beta)}}\right)_{i\in\mathbb{N}},$$

cf. Proposition 1(2). Furthermore, let $p \in [1, \infty]$.

- *If* $p > \frac{l+\beta}{l}$, *then:*
  - *(1)* $\zeta \in c_{||\cdot||_p}$ *almost surely since both*
    - *(a)* $||\zeta||_p < \infty$ *almost surely and*
    - *(b)* $\frac{D(n)}{n^{l/(l+\beta)}} \to \zeta$ almost surely in $c_{||\cdot||_p}$ *as* $n \to \infty$*;*
  - *(2)* $\sup_n \mathbb{E}\left(\left\| \frac{D(n)}{n^{l/(l+\beta)}} \right\|_p^{\alpha}\right) < \infty$ *for every* $\alpha > 0$.
- *If* $p < \frac{l+\beta}{l}$ *then:*
  - *(3)* $\frac{D(n)}{n^{l/(l+\beta)}} \not\to \zeta$ almost surely in $c_{||\cdot||_p}$ *as* $n \to \infty$.

The proof of this proposition is deferred to Appendix A.5.

Remark 4. Proposition 2 generalises the convergence results of Theorem 2 in [27] from weak to almost sure convergence and allows for general l and β.

4.3. Regular variation in sequence space

Theorem 6. *Consider a preferential attachment random graph with the notation as introduced in the last sections. Let* $D(n) := (D_1(n), D_2(n), \dots)$ *be the corresponding weight sequence at time* $n \in \mathbb{N}$ *and let* N *be a positive integer-valued random variable. Let* $p \in (\frac{l+\beta}{l}, \infty]$ *and* $||\cdot||$ *be a norm such that there exists some* $C > 0$ *with* $||\cdot|| \le C||\cdot||_p$. *If the conditions*

(1) N *and* $(D(n))_n$ *are independent and*

(2) N *is regularly varying with index* $\alpha > 0$

are satisfied, then $D(N)$ *is multivariate regularly varying in* $c_{||\cdot||}$ *with index* $\alpha \cdot \frac{l+\beta}{l}$.

Proof. Note first that our assumptions guarantee that the convergence results of Proposition 2 hold analogously with the norm $\|\cdot\|_p$ replaced by $\|\cdot\|$. Thus, the assumptions of Theorem 3 are met and the proof is completely analogous to the proof of Theorem 4.

This result finally allows us to derive (univariate) regular variation of the maximum degree of a preferential attachment model.

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

Corollary 3. *Under the assumptions of Theorem 6 the maximum degree* $\sup_{i\in\mathbb{N}} D_i(N)$ *is regularly varying with index* $\alpha \cdot \frac{l+\beta}{l}$.

Proof. This follows from Theorem 6 applied to $\|\cdot\|_\infty$ and Remark 2.

5. Conclusion

In this article, we proposed a novel approach to modelling large networks with power-law behaviour. Where previous limit results mainly focused on the limiting empirical degree distribution as the number of nodes goes to infinity, we provide a new perspective by stopping the preferential attachment process after a heavy-tailed number of steps, allowing us to consider the joint extremal behaviour of individual vertices. We proved multivariate regular variation under several model specifications, including the initial edge count per node and the offset parameter β, and analysed both the finite- and infinite-dimensional cases, where we identified the index of regular variation and characterised the spectral measure. From the limit results, it can be seen that the tail index of the individual nodes is influenced both by the tail index α of the number of steps, as well as by the parameters β and l of the preferential attachment model. In contrast, the extremal dependence structure depends on β and l, as well as the initial edge count. Taken together, these results provide a better understanding of the origins of power-law behaviour in random networks and enable us to estimate probabilities of extremal events.

Appendix A. Auxiliary results and deferred proofs

A.1. Proof of Proposition 1 and uniform integrability

For the proof of several lemmas and other auxiliary results we make use of a corollary to Stirling's formula, which can be found as Corollary 6.1.46 in [1] and reads as follows.

Lemma 5.

$$\lim_{n\to\infty} n^{b-a}\frac{\Gamma(a+n)}{\Gamma(b+n)} = 1 \quad \forall a, b \in \mathbb{R}.$$

This allows us to estimate binomial coefficients by their respective powers and *vice versa*.

Lemma 6. *Under the assumptions of Proposition 1, we have*

$$\prod_{i=1}^{r} \frac{D_i^{k_i}(n)}{\Gamma(k_i+1)} \sim \prod_{i=1}^{r} \binom{D_i(n)+k_i-1}{k_i} \quad \text{almost surely}$$

and there exist constants $C_1, C_2 > 0$ *independent of n such that almost surely*

$$C_1 \cdot \prod_{i=1}^{r} \binom{D_i(n)+k_i-1}{k_i} \le \prod_{i=1}^{r} D_i^{k_i}(n) \le C_2 \cdot \prod_{i=1}^{r} \binom{D_i(n)+k_i-1}{k_i}. \tag{A.1}$$

Proof. Remember from Lemma 2 that almost surely $D_i(n) \to \infty$ as $n \to \infty$. Thus, by Lemma 5, we can find, for each $i = 1, \dots, r$, constants $0 < C_1^i < C_2^i$ (depending on k_i) such that

$$C_1^i \binom{D_i(n)+k_i-1}{k_i} = C_1^i \frac{\Gamma(D_i(n)+k_i)}{\Gamma(k_i+1)\Gamma(D_i(n))} \le \frac{D_i^{k_i}(n)}{\Gamma(k_i+1)} \le C_2^i \binom{D_i(n)+k_i-1}{k_i}$$

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

holds for all $n \geq (r-s) \vee 0$. Combining those constants leads to (A.1).

Now to Proposition 1. The following proof adapts a martingale approach from [24] (Theorem 2.1).

Proof of Proposition 1.

(1) We call the total ball count at time m in the corresponding urn model

$$S_m := z + m + \beta \cdot \left\lfloor \frac{m}{l} \right\rfloor \quad \text{with } z := \sum_{i=1}^{N(0)} C_i(0) + j \cdot \beta.$$

Only one colour obtains a new ball at a time, so

$$\begin{aligned}
&\mathbb{E}\Big(\prod_{i=1}^{r} \binom{C_i(m+1)+k_i-1}{k_i} \Big| \mathcal{F}_m\Big) \\
&= \sum_{j=1}^{r} \mathbb{E}\Big(\prod_{\substack{i=1 \\ i\neq j}}^{r} \binom{C_i(m)+k_i-1}{k_i} \binom{C_j(m)+k_j}{k_j} \cdot \mathbb{1}_{\{C_j(m+1)=C_j(m)+1\}} \Big| \mathcal{F}_m\Big) \\
&\quad + \mathbb{E}\Big(\prod_{i=1}^{r} \binom{C_i(m)+k_i-1}{k_i} \cdot \mathbb{1}_{\bigcap_{j=1}^{r}\{C_j(m+1)=C_j(m)\}} \Big| \mathcal{F}_m\Big),
\end{aligned}$$

and using the transition probabilities of the urn model this equals

$$\begin{aligned}
&\sum_{j=1}^{r} \prod_{\substack{i=1 \\ i\neq j}}^{r} \binom{C_i(m)+k_i-1}{k_i} \cdot \binom{C_j(m)+k_j-1}{k_j} \cdot \frac{C_j(m)+k_j}{C_j(m)} \cdot \frac{C_j(m)}{S_m} \\
&+ \prod_{i=1}^{r} \binom{C_i(m)+k_i-1}{k_i} \cdot \Big(1 - \sum_{j=1}^{r} \frac{C_j(m)}{S_m}\Big) \\
&= \prod_{i=1}^{r} \binom{C_i(m)+k_i-1}{k_i} \Big(\sum_{j=1}^{r} \frac{C_j(m)+k_j}{S_m} + 1 - \sum_{j=1}^{r} \frac{C_j(m)}{S_m}\Big) \\
&= \prod_{i=1}^{r} \binom{C_i(m)+k_i-1}{k_i} \frac{S_m + k}{S_m}.
\end{aligned}$$

Now consider times $m = n \cdot l$. These are the critical times when a new colour is migrated into the urn (a new vertex is added to the graph). In between nothing of interest happens and we can iterate the above calculation l times to obtain

$$\mathbb{E}\Big(\prod_{i=1}^{r} \binom{C_i((n+1)\cdot l)+k_i-1}{k_i} \Big| \mathcal{F}_{n\cdot l}\Big) = \prod_{i=1}^{r} \binom{C(n\cdot l)+k_i-1}{k_i} \prod_{i=0}^{l-1} \frac{S_{n\cdot l+i}+k}{S_{n\cdot l+i}}.$$

In the next step we see that

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

$$\prod_{i=0}^{l-1}\frac{S_{n\cdot l+i}+k}{S_{n\cdot l+i}}=\prod_{i=0}^{l-1}\frac{z+n\cdot l+i+\beta\cdot n+k}{z+n\cdot l+i+\beta\cdot n}=\prod_{i=0}^{l-1}\frac{\frac{z+k+i}{l+\beta}+n}{\frac{z+i}{l+\beta}+n}$$
$$=\Big(\prod_{i=0}^{l-1}\frac{\Gamma(\frac{z+k+i}{l+\beta}+n+1)}{\Gamma(\frac{z+i}{l+\beta}+n+1)}\Big)\Big/\Big(\prod_{i=0}^{l-1}\frac{\Gamma(\frac{z+k+i}{l+\beta}+n)}{\Gamma(\frac{z+i}{l+\beta}+n)}\Big)=:\frac{c(n+1,k)}{c(n,k)},$$

and the martingale property follows. Lastly, we apply Lemma 5 to each factor of $c(n,k)$ to get

$$c(n,k)=\prod_{i=0}^{l-1}\frac{\Gamma(\frac{z+k+i}{l+\beta}+n)}{\Gamma(\frac{z+i}{l+\beta}+n)}\sim\prod_{i=0}^{l-1}n^{k/(l+\beta)}=n^{k\cdot l/(l+\beta)}. \tag{A.2}$$

(2) Doob's martingale convergence theorem (see Chapter XI.14 in [14]) guarantees the existence of an almost sure limit of the martingale given in (2.1) and that this limit is in $L^1(\mathbb{P})$. Lemma 6 in combination with (A.2) then ensures the almost sure convergence of the sequence $n^{-k_i l/(l+\beta)}D_i(n)^{k_i}$ to a limit $\zeta_i^{k_i}\in L^1(\mathbb{P})$, and a further application of Lemma 6 implies that the expression in (2.2) is indeed the limit of the process (2.1).

(3) Doob's martingale convergence theorem (Chapter XI.14 in [14]) yields that (2.2) is the right closure to the martingale (2.1) if the latter is uniformly integrable. To show this uniform integrability we show that the process is $L^{1+\epsilon}$-bounded for any $\epsilon>0$. To this end, set $p_i:=k_i(1+\epsilon)$ and $p:=k(1+\epsilon)=p_1+\ldots+p_r$. By Lemma 6 there exist constants $C_1, C_2>0$ independent of n such that

$$\Big[\frac{1}{c(n,k)}\prod_{i=1}^{r}\binom{D_i(n)+k_i-1}{k_i}\Big]^{1+\epsilon}\le C_1\cdot\frac{\prod_{i=1}^{r}D_i^{p_i}(n)}{n^{p\cdot l/(l+\beta)}}$$
$$\le C_2\cdot\frac{1}{c(n,p)}\prod_{i=1}^{r}\binom{D_i(n)+p_i-1}{p_i}.$$

Since the right-hand side is a multiple of a martingale, it has bounded expectation and the uniform integrability of (2.1), and thus the first part of (3), follows.
This now allows us to iteratively trace back the expectation of $\prod_{i=1}^{r}\zeta_i^{k_i}$ to the times at which colour i, $1\le i\le r$, was first introduced to the urn, that is, $n_0^i:=l(i-s)\vee 0$, and the number of balls was still deterministic, namely either β if $i>N(0)$ or $D_i(0)$ if $i\le N(0)$:

$$\mathbb{E}\Big(\prod_{i=1}^{r}\frac{\zeta_i^{k_i}}{\Gamma(k_i+1)}\Big)=\mathbb{E}\Big(\mathbb{E}\Big(\prod_{i=1}^{r}\frac{\zeta_i^{k_i}}{\Gamma(k_i+1)}\Big|\mathcal{F}_{n_0^r}\Big)\Big)$$
$$=\mathbb{E}\Big(\frac{1}{c(n_0^r/l,k_1+\ldots+k_r)}\prod_{i=1}^{r}\binom{D_i(n_0^r/l)+k_i-1}{k_i}\Big)$$

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

$$= \frac{c(n_0^r/l,\, k_1+\ldots+k_{r-1})}{c(n_0^r/l,\, k_1+\ldots+k_r)} \cdot \begin{cases} \binom{D_r(0)+k_r-1}{k_r} & \text{if } r \le N(0) \\ \binom{\beta+k_r-1}{k_r} & \text{if } r > N(0) \end{cases}$$

$$\cdot\, \mathbb{E}\Big(\mathbb{E}\Big(\frac{1}{c(n_0^r/l,\, k_1+\ldots+k_{r-1})} \prod_{i=1}^{r-1} \binom{D_i(n_0^r/l)+k_i-1}{k_i} \Big| \mathcal{F}_{n_0^{r-1}}\Big)\Big)$$

$$= \prod_{i=2}^{r} \frac{c(n_0^i/l,\, k_1+\ldots+k_{i-1})}{c(n_0^i/l,\, k_1+\ldots+k_i)} \cdot \frac{1}{c(n_0^1/l,\, k_1)} \prod_{i=1}^{r} \begin{cases} \binom{D_i(0)+k_i-1}{k_i} & \text{if } i \le N(0), \\ \binom{\beta+k_i-1}{k_i} & \text{if } i > N(0). \end{cases}$$

We end this section with a uniform integrability property that will turn out to be useful on multiple occasions.

Corollary 4. *Under the assumptions of and with the notation from Proposition 1 the process* $\big(n^{-\frac{k\cdot l}{l+\beta}} \prod_{i=1}^{r} D_i^{k_i}(n)\big)_n$ *is uniformly integrable.*

Proof. This follows from the uniform integrability of the martingale (2.1) shown in Proposition 1(1) and (3) together with Lemma 6.

A.2. Proof of Theorem 3

We adapt the proof from Theorem 3 in [37] to the case of a general separable Banach space. By our additional assumption about the continuity of (X_t) we are able to straighten a previously not properly addressed subtlety in the proof about the application of Egorov's theorem.

Proof of Theorem 3. By a Portmanteau theorem for $\mathbb{M}$-convergence, see Theorem 2.1 in [23], it is sufficient to show

$$\lim_{t\to\infty} t\mathbb{E}f\Big(\frac{X_T T}{b(t)}\Big) = \int_0^\infty \mathbb{E}f(X_\infty y)\, \nu_\alpha(\mathrm{d}y) < \infty, \tag{A.3}$$

for all non-negative, bounded and uniformly continuous functions f on B whose support is bounded away from 0, i.e. there exists an $\epsilon_0 > 0$ such that $f(x) = 0$ for all $x \in B_{\epsilon_0}(0)$. Without loss of generality, we will assume that f is bounded by 1. In order to show (A.3), we write, for $\eta, M > 0$,

$$\begin{aligned}
&\Big| t\mathbb{E}f\Big(\frac{X_T T}{b(t)}\Big) - \int_0^\infty \mathbb{E}f(X_\infty y)\nu_\alpha(\mathrm{d}y)\Big| \\
&\le \Big| t\mathbb{E}\Big(f\Big(\frac{X_T T}{b(t)}\Big)\mathbb{1}_{\{M\ge T/b(t)\ge\eta\}}\Big) - \int_\eta^M \mathbb{E}f(X_\infty y)\nu_\alpha(\mathrm{d}y)\Big| \\
&+ t\mathbb{E}\Big(f\Big(\frac{X_T T}{b(t)}\Big)\mathbb{1}_{\{T/b(t)<\eta\}}\Big) + \int_0^\eta \mathbb{E}f(X_\infty y)\nu_\alpha(\mathrm{d}y) \\
&+ t\mathbb{E}\Big(f\Big(\frac{X_T T}{b(t)}\Big)\mathbb{1}_{\{T/b(t)>M\}}\Big) + \int_M^\infty \mathbb{E}f(X_\infty y)\nu_\alpha(\mathrm{d}y) \\
&=: I^1(t,\eta,M) + I^2(t,\eta) + I^3(t,M).
\end{aligned}$$

(i). We start by showing $I^1(t,\eta,M) \to 0$, $t\to\infty$. To this end, first fix some $\epsilon > 0$. By Egorov's theorem there exists a measurable set A_ϵ with $\mathbb{P}(A_\epsilon^c) < \epsilon$ such that $X_t \to X_\infty$ for $t\to\infty$ uniformly on A_ϵ. At this point we use the one-sided continuity of (X_t) as a

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

sufficient condition to ensure the measurability of A_ϵ. This allows us to bound $I^1(t,\eta,M)$ separately on A_ϵ and A_ϵ^c:

$$\begin{aligned} I^1(t,\eta,M) &\le t\mathbb{E}\Big(f\Big(\frac{X_T T}{b(t)}\Big)\mathbb{1}_{A_\epsilon^c\cap\{M\ge T/b(t)\ge\eta\}}\Big) + \int_\eta^M \mathbb{E}\Big(f(X_\infty y)\mathbb{1}_{A_\epsilon^c}\Big)\nu_\alpha(\mathrm{d}y) \\ &\quad + t\mathbb{E}\Big(\Big|f\Big(\frac{X_T T}{b(t)}\Big) - f\Big(\frac{X_\infty T}{b(t)}\Big)\Big|\mathbb{1}_{A_\epsilon\cap\{M\ge T/b(t)\ge\eta\}}\Big) \\ &\quad + \Big|t\mathbb{E}\Big(f\Big(\frac{X_\infty T}{b(t)}\Big)\mathbb{1}_{A_\epsilon\cap\{M\ge T/b(t)\ge\eta\}}\Big) - \int_\eta^M \mathbb{E}\Big(f(X_\infty y)\mathbb{1}_{A_\epsilon}\Big)\nu_\alpha(\mathrm{d}y)\Big| \\ &=: I_a^1(t,\eta,M) + I_b^1(t,\eta,M) + I_c^1(t,\eta,M) + I_d^1(t,\eta,M). \end{aligned}$$

Starting with $I_a^1(t,\eta,M)$, we use that $A_\epsilon^c \in \sigma(X_t, t\ge 0)$ which therefore is independent of T:

$$\limsup_{t\to\infty} I_a^1(t,\eta,M) \le \limsup_{t\to\infty} t\mathbb{P}\Big(A_\epsilon^c \cap \Big\{\frac{T}{b(t)}\ge\eta\Big\}\Big) \le \epsilon\cdot\eta^{-\alpha}.$$

Second,

$$I_b^1(t,\eta,M) \le \mathbb{P}(A_\epsilon^c)\int_\eta^M \nu_\alpha(\mathrm{d}y) \le \epsilon\cdot(\eta^{-\alpha} - M^{-\alpha}).$$

For $I_c^1(t,\eta,M)$, we note that $b(t)\to\infty$ for $t\to\infty$, yielding an arbitrarily large lower bound for T on the set $\{T/b(t)\ge\eta\}$. Then, intersecting with A_ϵ we obtain that $b(t)^{-1}T||X_T - X_\infty||\cdot\mathbb{1}_{A_\epsilon\cap\{M\ge T/b(t)\ge\eta\}}$ uniformly tends to 0 and combining this with the uniform continuity of f we find a $c(t)$ with $c(t)\to 0$ for $t\to\infty$ such that, for $t\to\infty$,

$$I_c^1(t,\eta,M) \le c(t)\cdot t\mathbb{P}\Big(A_\epsilon\cap\Big\{M\ge\frac{T}{b(t)}\ge\eta\Big\}\Big) \le c(t)\cdot t\mathbb{P}\Big(\frac{T}{b(t)}\ge\eta\Big)\to 0.$$

Finally, for $I_d^1(t,\eta,M)$ we observe that

$$y\mapsto E(f(X_\infty y)\mathbb{1}_{A_\epsilon})\mathbb{1}_{[\eta,M]}(y)$$

is bounded and has support bounded away from 0. Even though it is not continuous in y, a standard approximation argument by continuous functions combined with dominated convergence gives

$$\begin{aligned} &t\mathbb{E}\Big(f\Big(\frac{X_\infty T}{b(t)}\Big)\mathbb{1}_{A_\epsilon\cap\{M\ge T/b(t)\ge\eta\}}\Big) \\ &= \int_\eta^M \mathbb{E}(f(X_\infty y)\mathbb{1}_{A_\epsilon})t\mathbb{P}^{T/b(t)}(\mathrm{d}y) \to \int_\eta^M \mathbb{E}(f(X_\infty y)\mathbb{1}_{A_\epsilon})\nu_\alpha(\mathrm{d}y) \end{aligned}$$

for $t\to\infty$, where we used regular variation of T in combination with Fubini's theorem, as T is independent of X_∞. Thus, $I_d(t,\eta,M)\to 0$ for $t\to\infty$. From the above and since we can choose $\epsilon>0$ arbitrarily small, we see that $I^1(t,\eta,M)\to 0$, $t\to\infty$.

(ii). We split up $I^2(t,\eta)$ into

$$t\mathbb{E}\Big(f\Big(\frac{X_T T}{b(t)}\Big)\mathbb{1}_{\{T/b(t)<\eta\}}\Big) + \int_0^\eta \mathbb{E}f(X_\infty y)\nu_\alpha(\mathrm{d}y) =: I_a^2(t,\eta) + I_b^2(\eta).$$

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

Then, using Markov's inequality as well as the independence of $(X_t)_{t\geq 0}$ and T, we arrive at the upper bound

$$\begin{aligned}
I_a^2(t,\eta) &\leq t\mathbb{P}\Big(\Big\|\frac{X_T T}{b(t)}\Big\| > \epsilon_0,\ \frac{T}{b(t)} < \eta\Big)\\
&= t\mathbb{P}\Big(\Big\|\frac{X_T T}{b(t)}\mathbb{1}_{\{T/b(t)<\eta\}}\Big\| > \epsilon_0\Big) \leq \epsilon_0^{-\alpha'} t\mathbb{E}\Big(\Big\|\frac{X_T T}{b(t)}\mathbb{1}_{\{T/b(t)<\eta\}}\Big\|^{\alpha'}\Big)\\
&= \epsilon_0^{-\alpha'}\int_{(0,\eta)} \mathbb{E}||X_{b(t)y}||^{\alpha'} y^{\alpha'} t\mathbb{P}(T/b(t)\in \mathrm{d}y)\\
&\leq \epsilon_0^{-\alpha'} \sup_{t\geq 0}\mathbb{E}||X_t||^{\alpha'} t\mathbb{E}\Big(\Big(\frac{T}{b(t)}\Big)^{\alpha'}\mathbb{1}_{\{T/b(t)<\eta\}}\Big).
\end{aligned}$$

By Karamata's theorem applied to truncated moments of regularly varying random variables, see Proposition 1.4.6 in [21], we get

$$\begin{aligned}
&t\mathbb{E}\Big(\Big(\frac{T}{b(t)}\Big)^{\alpha'}\mathbb{1}_{\{T/b(t)<\eta\}}\Big)\\
&= t\mathbb{P}(T>b(t))\frac{\mathbb{E}\Big(T^{\alpha'}\mathbb{1}_{\{T<\eta b(t)\}}\Big)}{(b(t))^{\alpha'}\mathbb{P}(T>b(t))}\\
&\to \frac{\alpha}{\alpha'-\alpha}\eta^{\alpha'-\alpha}.
\end{aligned}$$

Thus, for $t\to\infty$, $\eta\to 0$, $I_a^2(t,\eta)\to 0$.

In order to bound $I_b^2(\eta)$, we first note that $\mathbb{E}||X_\infty||^\beta$ is finite for all $\alpha<\beta<\alpha'$ because $(||X_t||^\beta)_t$ is uniformly integrable (as $\sup_t \mathbb{E}\big[(||X_t||^\beta)^{\frac{\alpha'}{\beta}}\big]<\infty$ and $\frac{\alpha'}{\beta}>1$) and the Vitali convergence theorem implies that $||X_t||^\beta$ converges in L^1 to $||X_\infty||^\beta$ and so do the expected values. Then, again by Markov's inequality, we get

$$\begin{aligned}
&\int_0^\eta \mathbb{E}f(X_\infty y)\nu_\alpha(\mathrm{d}y) \leq \int_0^\eta \mathbb{P}(\|X_\infty\|>\epsilon_0/y)\nu_\alpha(\mathrm{d}y)\\
&\leq \int_0^\eta \mathbb{E}\big(\|X_\infty\|^\beta\big)\epsilon_0^{-\beta}y^\beta\nu_\alpha(\mathrm{d}y)\\
&\leq \mathbb{E}\big(\|X_\infty\|^\beta\big)\epsilon_0^{-\beta}\frac{\alpha}{\beta-\alpha}\eta^{\beta-\alpha}.
\end{aligned}$$

Again, with $\eta\to 0$ we have shown that $I_b^2(\eta)\to 0$ and therefore $I^2(t,\eta)\to 0$, and also that the right-hand side in (A.3) is finite, since $\int_\eta^\infty \mathbb{E}f(X_\infty y)\nu_\alpha(\mathrm{d}y)\leq \eta^{-\alpha}$.

(iii). Finally, regarding $I^3(t,M)$ we note that

$$\limsup_{t\to\infty} I^3(t,M) \leq \limsup_{t\to\infty} t\mathbb{P}(T>Mb(t)) + \int_M^\infty \nu_\alpha(dy) \leq 2M^{-\alpha},$$

and so $\lim_{M\to\infty}\limsup_{t\to\infty} I^3(t,M)=0$. This finishes the proof of (A.3).

Remark 5. One could do without the continuity assumption of (X_t) provided the measurability of all sets in the proof is ensured. In particular, this refers to the A_ϵ from Egorov's theorem.

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

A.3. *Moment conditions* The following statement extends Lemma 4.1 in [26] to arbitrary values of $\beta \geq 0$.

Lemma 7. *Assume the urn model of Section 2.1.2, fix $k \in \mathbb{N}$ and let*

$$U_k(n) := C_1(n) + \ldots + C_k(n)$$

denote the total number of balls of colours 1 to k at time n. Then for every $q \in \mathbb{N}$ there exist $c, C > 0$ such that, for all $n \in \mathbb{N}_0$,

$$cn^{ql/(l+\beta)} < \mathbb{E}(U_k(n)^q) < Cn^{ql/(l+\beta)}. \tag{A.4}$$

Proof. Let $k \in \mathbb{N}$ be fixed throughout this proof. Note first that $U_k(n)$, $n \in \mathbb{N}_0$, is deterministic until there are at least $k+1$ different colours in the urn, i.e. for $n \leq n_0$ with

$$n_0 := l(k+1-s) \vee 0,$$

and set

$$n_p := n_0 + p, \ p \in \mathbb{N}.$$

It is then sufficient to show (A.4) for $n = n_p$, $p \in \mathbb{N}$, as for those finitely many $n \leq n_0$ there exist trivial bounds. Let $h_p = U_{(k+1)\vee s}(n_0) + p + \beta\lfloor \frac{p}{l} \rfloor$ be the (deterministic) total number of balls at time n_p and define

$$M_{p,q} := \prod_{i=0}^{q-1} (i + U_k(n_p)), \ p \in \mathbb{N}_0, \ q \in \mathbb{N}. \tag{A.5}$$

We start the proof by showing that

$$\mathbb{E}M_{p,q} = M_{0,q} \cdot \prod_{i=0}^{p-1} \Big(1 + \frac{q}{h_i}\Big). \tag{A.6}$$

Given the value $U_k(n_{p-1})$, there are only two possible values for $U_k(n_p)$, which are $U_k(n_{p-1}) + 1$ and $U_k(n_{p-1})$: either we draw a ball of colours 1 to k or we do not. Using this, we get

$$\begin{aligned}
\mathbb{E}(M_{p,q}|U_k(n_{p-1})) &= \frac{U_k(n_{p-1})}{h_{p-1}} \prod_{i=0}^{q-1} (i + U_k(n_{p-1}) + 1) + \frac{h_{p-1} - U_k(n_{p-1})}{h_{p-1}} M_{p-1,q} \\
&= \frac{U_k(n_{p-1})}{h_{p-1}} \prod_{i=1}^{q} (i + U_k(n_{p-1})) + \frac{h_{p-1} - U_k(n_{p-1})}{h_{p-1}} M_{p-1,q} \\
&= \frac{U_k(n_{p-1})}{h_{p-1}} \frac{M_{p-1,q}(U_k(n_{p-1}) + q)}{U_k(n_{p-1})} + \frac{h_{p-1} - U_k(n_{p-1})}{h_{p-1}} M_{p-1,q} \\
&= M_{p-1,q}\Big(1 + \frac{q}{h_{p-1}}\Big).
\end{aligned}$$

By iterating this calculation p times, one obtains (A.6).

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

Now, by definition of h_p, $p \in \mathbb{N}$, we have

$$U_{(k+1)\vee s}(n_0) + p + \beta\Big(\frac{p}{l} - 1\Big) \le h_p \le U_{(k+1)\vee s}(n_0) + p + \beta\frac{p}{l}.$$

Setting $x := \frac{l}{l+\beta}$ and $y := (U_{(k+1)\vee s}(n_0) - \beta)\frac{l}{l+\beta}$ (which is non-negative as for each colour there are at least β balls) for brevity, this is equivalent to

$$\frac{y+p}{x} \le h_p \le \frac{y+p+\beta x}{x}. \tag{A.7}$$

For the upper bound in (A.4) use (A.6) and (A.7) to get

$$\mathbb{E}M_{p,q} \le M_{0,q} \cdot \prod_{i=0}^{p-1}\Big(1 + \frac{qx}{y+i}\Big) = M_{0,q} \cdot \frac{\Gamma(y)}{\Gamma(qx+y)} \cdot \frac{\Gamma(qx+y+p)}{\Gamma(y+p)}.$$

Thus, using Lemma 5, there exist $\tilde{C}$, $C > 0$ such that

$$\begin{aligned}\mathbb{E}U_k(n_p)^q \le \mathbb{E}M_{p,q} &\le M_{0,q} \cdot \frac{\Gamma(y)}{\Gamma(qx+y)} \cdot \frac{\Gamma(qx+y+p)}{\Gamma(y+p)} \\ &\le \tilde{C}(y+p)^{qx} < Cn_p^{qx} = Cn_p^{ql/(l+\beta)}\end{aligned}$$

for all $p \in \mathbb{N}$, proving the upper bound.

For the lower bound, use Jensen's inequality, (A.6), (A.7), Lemma 5 and similar reasoning as above to see that there exist $\tilde{c}$, $c > 0$ such that

$$\begin{aligned}\mathbb{E}U_k(n_p)^q = \mathbb{E}M_{p,1}^q &\ge (\mathbb{E}M_{p,1})^q \\ &\ge \left(M_{0,1} \cdot \prod_{i=0}^{p-1}\Big(1 + \frac{x}{y+i+\beta x}\Big)\right)^q \\ &= M_{0,1}^q \left(\frac{\Gamma(y+\beta x)}{\Gamma(x+y+\beta x)}\right)^q \left(\frac{\Gamma(x+y+\beta x+p)}{\Gamma(y+\beta x+p)}\right)^q \\ &\ge \tilde{c}(y+\beta x+p)^{qx} > cn_p^{qx} = cn_p^{ql/(l+\beta)},\end{aligned}$$

for all $p \in \mathbb{N}$, which finishes the proof.

A.4. Proof of Lemma 4

(1) We adapt the proof of Lemma 3.3 from [24]. By definition, $B_k^{\downarrow} = \zeta_k/(\zeta_1 + \ldots + \zeta_k)$ and so the statement follows immediately for $k = 1$. Thus, keep $k > 1$ fixed in the following. We disregard all colours in the urn model but 1 through k and further simplify the model. To find the distribution of $B_k^{\downarrow}$ it is sufficient to view all balls of colours 1 to $k-1$ as being 'black' and those of colour k as being 'white'. At the time n_0^k, when the last relevant colour is added to the urn, the number w_k of white and b_k of black balls is deterministic and given by

$$(w_k, b_k) = \begin{cases} (C_k(0), \sum_{i=1}^{k-1} C_i(0)) & \text{if } 1 < k \le s, \\ (\beta, \sum_{i=1}^{s} C_i(0) + (l+\beta)(k-1-s) + l) & \text{if } k > s. \end{cases}$$

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

Now at each time either the number of black and white balls does not change or a new ball is added to one of them. Conditionally on the latter, the transition probabilities are the same as in a traditional Pólya urn and therefore Lemma 3 yields that the vector $(B_k^{\downarrow}, 1 - B_k^{\downarrow})$ has a Dirichlet distribution, whose marginals are known to be beta distributions with the given parameters.

(2) We provide an alternative construction for $C^r(n)$ to help us derive the desired independence properties. We start with the usual setup of an urn with s different colours and starting amounts $C_0(0), \ldots, C_s(0)$. From there on, instead of picking a ball directly, in each time step we flip a sequence of independent Bernoulli coins, determined below, until we observe heads for the first time. Then we add a new ball of one colour, based on which flip resulted in heads, and proceed to the next time step. The individual coins may have different success probabilities, i.e. probabilities of showing heads. More precisely, for the ball to be selected at time n, let m be the currently largest numbered colour in the urn. If $m > r$ the first flip in a time step always represents the colours $> r$. If that coin flip shows heads we add a new ball to one of those (which one specifically is irrelevant to us). If it shows tails, however, we start traversing the remaining r colours in backwards order. This means we start by flipping for colour r, then either add a new ball to it if the flip showed heads or otherwise continue to $r-1$ and so on. If $m \leq r$ we skip the flips for colours $m+1, m+2, \ldots, `> r'$. To also incorporate the migration of new colours in this model, after every lth time step, we add additional β balls to colour $m+1$ or if $m \geq r$ to the '$> r$' category. Next, for this process to have the same distribution as $(C^r(n))_{n\in\mathbb{N}_0}$, we need to find suitable success probabilities $p_i(n)$ when flipping for colour i in time step $n \mapsto n+1$. To this end, we write

$$U_k(n) := \sum_{i=1}^{k} C_i(n), \quad k \geq 1,$$

for brevity and set

$$\text{if } i \leq r: \qquad p_i(n) = \frac{C_i(n)}{C_1(n) + \ldots + C_i(n)} = 1 - \frac{U_{i-1}(n)}{U_i(n)},$$

$$\text{and} \qquad p_{>r}(n) = \frac{C_{r+1}(n) + \ldots + C_m(n)}{C_1(n) + \ldots + C_m(n)} = 1 - \frac{U_r(n)}{U_m(n)}.$$

To verify that this yields exactly the same transition probabilities as in the urn model, write (with $\mathcal{F}_n = \sigma(C_i(t) : t \leq n,\ i, t \in \mathbb{N})$) for colours $i \leq r$,

$$\begin{aligned}\mathbb{P}(C_i(n+1) - C_i(n) = 1 | \mathcal{F}_n) &= \frac{C_i(n)}{U_m(n)} = \frac{U_r(n)}{U_m(n)} \frac{U_{r-1}(n)}{U_r(n)} \cdots \frac{U_i(n)}{U_{i+1}(n)} \frac{C_i(n)}{U_i(n)} \\ &= (1 - p_{>r}(n)) \cdot (1 - p_r(n)) \cdot \ldots \cdot (1 - p_{i+1}(n)) \cdot p_i(n)\end{aligned}$$

and for colour '$> r$',

$$\mathbb{P}([U_m(n+1) - U_r(n+1)] - [U_m(n) - U_r(n)] = 1 | \mathcal{F}_n) = \frac{U_m(n) - U_r(n)}{U_m(n)} = p_{>r}(n).$$

So we conclude that the two processes indeed have the same transition probabilities and proceed to show the independence of the $B_i^{\downarrow}$.

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

For that let $(Y_j^i)_{j=1}^\infty$ be the random variables which contain the outcome of the jth flip for colour i, $i = 1, \ldots, r$. Define T_j^i to be the step in the above urn model when coin Y_j^i is flipped, noting that T_j^i, $j > 0$, is random but almost surely finite by Lemma 2. The success probability $p_i(n)$ at time $n = T_j^i$ is dependent on prior flips and given by

$$\frac{C_i(n)}{C_1(n) + \ldots + C_i(n)} = \frac{C_i(T_1^i) + \sum_{k=1}^{j-1} Y_k^i}{U_i(T_1^i) + j - 1}$$

for colours $i \le r$, as $U_i(T_1^i)$ balls of colours 1 to i and $C_i(T_1^i)$ balls of colour i are present when we first flip the coin for this colour, and at the jth flip a further $j-1$ balls of colours 1 to i have already been added to the urn, with $\sum_{k=1}^{j-1} Y_k^i$ of them to colour i. Therefore, the $p_i(n)$ at time $n = T_j^i$ are completely determined by deterministic starting values $U_i(T_1^i)$, $C_i(T_1^i)$ and the previous flips for colour i and the sequences $(Y_j^1)_j, \ldots, (Y_j^r)_j$ are jointly independent.

With this, the independence of the $B_i^\downarrow$ follows since, for $j \to \infty$,

$$\frac{C_i(T_1^i) + \sum_{k=1}^{j-1} Y_k^i}{U_i(T_1^i) + j - 1} = \frac{C_i(T_j^i)}{C_1(T_j^i) + \ldots + C_i(T_j^i)} \to B_i^\downarrow \qquad \text{for } i = 2, \ldots, r.$$

(3) In addition to the sequences introduced in (2), let $(Y_j^{>r})_{j=1}^\infty$ be the random variables which contain the outcome of the jth flip for colour '$> r$' and let $T_j^{>r}$, $j > 0$, be the step in the urn model when coin $Y_j^{>r}$ is flipped. The first flip $T_1^{>r}$ takes place at time n_0^{r+1} with $U_{(r+1)\vee s}(n_0^{r+1})$ total balls in the urn, and with $U_{(r+1)\vee s}(n_0^{r+1}) - U_r(n_0^{r+1})$ of them of colour '$> r$', both numbers being deterministic. After n_0^{r+1}, the coin for colour '$> r$' is flipped in every consecutive step of the urn model. Thus, at the jth flip for this colour, the total number of balls has grown to $U_{(r+1)\vee s}(n_0^{r+1}) + j - 1 + \beta\lfloor\frac{j-1}{l}\rfloor$ (where the last summand is due to the immigration), and the number of colour '$> r$' has grown to $U_{(r+1)\vee s}(n_0^{r+1}) - U_r(n_0^{r+1}) + \sum_{k=1}^{j-1} Y_k^{>r} + \beta\lfloor\frac{j-1}{l}\rfloor$. Therefore, the success probability $p_{>r}(n)$ at time $n = T_j^{>r}$ is given by

$$\frac{U_{(r+1)\vee s}(n_0^{r+1}) - U_r(n_0^{r+1}) + \sum_{k=1}^{j-1} Y_k^{>r} + \beta\lfloor\frac{j-1}{l}\rfloor}{U_{(r+1)\vee s}(n_0^{r+1}) + j - 1 + \beta\lfloor\frac{j-1}{l}\rfloor},$$

and the sequence $(Y_j^{r>})_{j=1}^\infty$ is independent of the sequences introduced in (2).

Observe that

$$\frac{1}{j^{l/(l+\beta)}} U_r(n_0^{r+1} + j) = \frac{1}{j^{l/(l+\beta)}} (U_r(n_0^{r+1}) + \sum_{k=1}^{j} (1 - Y_k^{>r})) \overset{j\to\infty}{\to} \zeta_1 + \ldots + \zeta_r,$$

and by independence of the coin flip sequences the limit $\zeta_1 + \ldots + \zeta_r$ is thus independent of $(B_1^\downarrow, \ldots, B_r^\downarrow)$ and thus independent of $(B_1, \ldots, B_r)$ as well.

A.5. Proof of Proposition 2

We start by showing (2).

(2) Case $\alpha \le p$: Let $\alpha' > p \ge \alpha$, then by Hölder's inequality

$$\sup_n \mathbb{E}\Big(\Big\| \frac{D(n)}{n^{l/(l+\beta)}} \Big\|_p^\alpha\Big) \le \sup_n \Big(\mathbb{E}\Big(\Big\| \frac{D(n)}{n^{l/(l+\beta)}} \Big\|_p^{\alpha'}\Big)\Big)^{\frac{\alpha}{\alpha'}} = \Big(\sup_n \mathbb{E}\Big(\Big\| \frac{D(n)}{n^{l/(l+\beta)}} \Big\|_p^{\alpha'}\Big)\Big)^{\frac{\alpha}{\alpha'}}$$

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

and this case follows whenever there is a bound for the next one.

Case $\alpha > p$: Because of the inverse Minkowski's inequality (see, e.g., III 2.4 Theorem 9 in [8]), p-quasinorms with $0 < p < 1$ are concave on $\mathbb{R}^d_{\geq 0}$. We can thus apply Jensen's inequality to

$$\mathbb{E}\Big(\Big\|\frac{D(n)}{n^{l/(l+\beta)}}\Big\|_p^\alpha\Big) = \mathbb{E}\Big(\Big\|\Big(\frac{D(n)}{n^{l/(l+\beta)}}\Big)^\alpha\Big\|_{\frac{p}{\alpha}}\Big) \leq \Big(\sum_{i=1}^{N(n)}\Big(\mathbb{E}\frac{D_i^\alpha(n)}{n^{\alpha\cdot l/(l+\beta)}}\Big)^{\frac{p}{\alpha}}\Big)^{\frac{\alpha}{p}}.$$

An application of Proposition 1(1) and Lemma 6 yields that there exists a $C > 0$ such that the right-hand side is bounded above by

$$C\cdot\left(\sum_{i=1}^{N(n)}\left(c(n,\alpha)^{-1}\mathbb{E}\begin{pmatrix}D_i(n)+\alpha-1\\ \alpha\end{pmatrix}\right)^{\frac{p}{\alpha}}\right)^{\frac{\alpha}{p}}.$$

Now, each base in the summands on the right is the expectation of a closed martingale and by Proposition 1(3),

$$\left(\sum_{i=1}^{N(n)}\left(c(n,\alpha)^{-1}\mathbb{E}\begin{pmatrix}D_i(n)+\alpha-1\\ \alpha\end{pmatrix}\right)^{\frac{p}{\alpha}}\right)^{\frac{\alpha}{p}} = \Big(\sum_{i=1}^{N(n)}(\mathbb{E}\zeta_i^\alpha)^{\frac{p}{\alpha}}\Big)^{\frac{\alpha}{p}}. \tag{A.8}$$

By (2.3) and Proposition 1(1) we have, for any fixed $k \in \mathbb{R}_{\geq 0}$ and all $i > N(0)$,

$$\mathbb{E}(\zeta_i^k) = C(k)c((i-s),k)^{-1} \sim C(k)(i-s)^{-k\cdot l/(l+\beta)}, \tag{A.9}$$

where $C(k)$ denotes a constant that does not depend on i. Thus, the limit as $n\to\infty$ of the right-hand side of (A.8) is finite if and only if

$$\frac{\alpha\cdot l}{l+\beta}\cdot\frac{p}{\alpha} > 1 \quad\Leftrightarrow\quad p > \frac{l+\beta}{l},$$

which is guaranteed by our assumption.

(1) (a) Assume first that $\frac{l+\beta}{l} < p < \infty$. Use monotone convergence to get

$$\mathbb{E}(\|\zeta\|_p^p) = \mathbb{E}\sum_{i=1}^\infty \zeta_i^p = \sum_{i=1}^\infty \mathbb{E}\zeta_i^p. \tag{A.10}$$

Again from (A.9), this sum is finite under our assumption that $p > (1+\beta)/l$ and so $\|\zeta\|_p^p < \infty$ almost surely. Furthermore, since $\|\zeta\|_\infty \leq \|\zeta\|_p$, the result also follows for $p = \infty$.

(1) (b) Next we prove the almost sure convergence in $c_{||\cdot||_p}$. We need to show that

$$\Big\|\frac{D(n)}{n^{l/(l+\beta)}} - \zeta\Big\|_p^p = \sum_{i=1}^\infty\Big|\frac{D_i(n)}{n^{l/(l+\beta)}} - \zeta_i\Big|^p \to 0,\ n\to\infty \quad \text{almost surely.}$$

By Scheffé's theorem the above convergence follows from componentwise convergence, i.e.

$$\frac{D_i(n)}{n^{l/(l+\beta)}} \to \zeta_i,\ n\to\infty \quad \text{almost surely,} \tag{A.11}$$

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

for all $i \in \mathbb{N}$ in combination with convergence of the norms, which is

$$\sum_{i=1}^{N(n)} \frac{D_i(n)^p}{n^{pl/(l+\beta)}} \to \sum_{i=1}^{\infty} \zeta_i^p, \ n \to \infty. \tag{A.12}$$

Proposition 1(2) implies (A.11), so we are left to show (A.12). To this end, introduce the random variables $X_i(n) := c(n,p)^{-1} \binom{D_i(n)+p-1}{p}$, which form, for each $i \in \mathbb{N}$, a martingale, according to Proposition 1(1). Their cumulative sums form submartingales, since

$$\mathbb{E}\Big(\sum_{i=1}^{N(n)} X_i(n) \Big| \mathcal{F}_{(n-1)l}\Big) = \sum_{i=1}^{N(n)} \mathbb{E}(X_i(n)|\mathcal{F}_{(n-1)l}) \geq \sum_{i=1}^{N(n-1)} X_i(n-1), \ i \in \mathbb{N}.$$

By Lemma 6 there exists, for each $k \in \mathbb{N}$, a C such that

$$\mathbb{E}\Big(\Big(\sum_{i=1}^{N(n)} X_i(n)\Big)^k\Big) \leq C\mathbb{E}\Big(\sum_{i=1}^{N(n)} \frac{D_i(n)^p}{n^{pl/(l+\beta)}}\Big)^k.$$

By (2) with $\alpha = pk$, we observe that the right-hand side is bounded uniformly in n. This means the submartingale is uniformly integrable and therefore convergent almost surely and in L_1 (Chapter XI.14 in [14]). To derive the limit we first note that, by Proposition 1(2),

$$\lim_{n\to\infty} \sum_{i=1}^{N(n)} X_i(n) - \sum_{i=1}^{l} \frac{\zeta_i^p}{\Gamma(p+1)} = \lim_{n\to\infty} \sum_{i=l+1}^{N(n)} X_i(n) \geq 0$$

for all $l \in \mathbb{N}$. Let $l \to \infty$ to conclude that

$$\lim_{n\to\infty} \sum_{i=1}^{N(n)} X_i(n) - \sum_{i=1}^{\infty} \frac{\zeta_i^p}{\Gamma(p+1)} \geq 0.$$

This implies that

$$\begin{aligned} &E\Big(\Big|\lim_{n\to\infty} \sum_{i=1}^{N(n)} X_i(n) - \sum_{i=1}^{\infty} \frac{\zeta_i^p}{\Gamma(p+1)}\Big|\Big) \\ &= E\Big(\lim_{n\to\infty} \sum_{i=1}^{N(n)} X_i(n) - \sum_{i=1}^{\infty} \frac{\zeta_i^p}{\Gamma(p+1)}\Big) \\ &= \lim_{n\to\infty} \sum_{i=1}^{N(n)} E(X_i(n)) - \sum_{i=1}^{\infty} E\Big(\frac{\zeta_i^p}{\Gamma(p+1)}\Big) = 0, \end{aligned}$$

where we used the L_1-convergence of our submartingale together with (A.10) in the penultimate step and Proposition 1(2) in the last step. Thus,

$$\lim_{n\to\infty} \sum_{i=1}^{N(n)} c(n,p)^{-1} \binom{D_i(n)+p-1}{p} = \sum_{i=1}^{\infty} \frac{\zeta_i^p}{\Gamma(p+1)} \text{ almost surely.} \tag{A.13}$$

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

Returning to the original process, we can use Proposition 1(1) and Lemma 6 to find a $C > 0$ such that

$$\frac{D_i(n)^p}{n^{pl/(l+\beta)}} \leq C \cdot c(n,k)^{-1} \binom{D_i(n)+p-1}{p}$$

for all $i = 1, \ldots, N(n)$, and thus apply Pratt's lemma [29] to finally arrive at (A.12) and therefore conclude (1)(b).

(3) For $p \in (1, \infty)$ the p-norm on $\mathbb{R}^d$ among all vectors of equal ℓ_1-norm is minimised if each component takes the same value. Let c be the cumulated starting weight of the initial $N(0)$ vertices. Then

$$\left\| \frac{D(n)}{n^{l/(l+\beta)}} \right\|_p^p = \frac{1}{n^{p\cdot l/(l+\beta)}} \sum_{i=1}^{N(n)} D_i^p(n) \geq \frac{(N(0)+n)\cdot(\frac{c+n(l+\beta)}{N(0)+n})^p}{n^{p\cdot l/(l+\beta)}} \sim C \cdot n^{1-\frac{p\cdot l}{l+\beta}}$$

for some $C > 0$ and, since by assumption $p < \frac{l+\beta}{l}$, the right-hand side diverges to ∞. Being almost surely unbounded, the sequence $D(n)/n^{l/(l+\beta)}$ thus fails to converge.

Acknowledgements

We thank the two anonymous referees and the associate editor for their helpful feedback on the manuscript.

Funding Information

There are no funding bodies to thank relating to the creation of this article.

Competing Interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] Abramowitz, M. and Stegun, I. A. (eds) (1984). *Pocketbook of Mathematical Functions*, abridged edn of Abramowitz/Stegun (eds) Handbook of Mathematical Functions, 1964 edn. Deutsch, Thun; Frankfurt am Main.

[2] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science* **286**, 509–512.

[3] Berger, N., Borgs, C., Chayes, J. T. and Saberi, A. (2005). On the spread of viruses on the internet. In Proc. 16th ACM–SIAM Symposium on Discrete Algorithm (SODA). pp. 301–310.

[4] Berger, N., Borgs, C., Chayes, J. T. and Saberi, A. (2014). Asymptotic behavior and distributional limits of preferential attachment graphs. *The Annals of Probability* 1–40.

[5] Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via pólya urn schemes. *The Annals of Statistics* **1**, 353–355.

[6] Bollobás, B., Borgs, C., Chayes, J. T. and Riordan, O. (2003). Directed scale-free graphs. In Proc. 14th Annual ACM–SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003). ACM, New York, pp. 132–139.

[7] Bollobás, B., Riordan, O., Spencer, J. and Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures & Algorithms* **18**, 279–290.

[8] Bullen, P. S. (2003). *Handbook of Means and Their Inequalities* (Mathematics and Its Applications), 2nd edn. Springer, Dordrecht.

[9] Collevecchio, A., Cotar, C. and LiCalzi, M. (2013). On a preferential attachment and generalized Pólya's urn model. *The Annals of Applied Probability* 1219–1253.

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press

[10] CONNOR, R. J. AND MOSIMANN, J. E. (1969). Concepts of independence for proportions with a generalization of the dirichlet distribution. *Journal of the American Statistical Association* **64**, 194–206.
[11] DE HAAN, L. AND FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*, Springer.
[12] DE HAAN, L. AND RESNICK, S. I. (1979). Derivatives of regularly varying functions in $\mathbb{R}^d$ and domains of attraction of stable distributions. *Stochastic Processes and their Applications* **8**, 349–355.
[13] DEIJFEN, M., VAN DEN ESKER, H., VAN DER HOFSTAD, R. AND HOOGHIEMSTRA, G. (2009). A preferential attachment model with random initial degrees. *Arkiv för matematik* **47**, 41–72.
[14] DOOB, J. L. (1994). *Measure Theory* (Graduate Texts in Mathematics, 143). Springer, New York.
[15] GARAVAGLIA, A. AND STEGEHUIS, C. (2019). Subgraphs in preferential attachment models. *Advances in Applied Probability* **51**, 898–926.
[16] HULT, H. AND LINDSKOG, F. (2006). Regular variation for measures on metric spaces. *Publications de l'Institut Mathématique* **80**, 121–140.
[17] JANSON, S. AND WARNKE, L. (2021). Preferential attachment without vertex growth: emergence of the giant component. *The Annals of Applied Probability* **31**, 1523–1547.
[18] JESSEN, H. A. AND MIKOSCH, T. (2006). Regularly varying functions. *Publications de L'institut Mathematique* **80**, 171–192.
[19] KIFER, Y. (1986). *Ergodic Theory of Random Transformations*. Birkhäuser, Boston, MA.
[20] KRAPIVSKY, P. L., RODGERS, G. J. AND REDNER, S. (2001). Degree distributions of growing networks. *Physical Review Letters* **86**, 5401–5404.
[21] KULIK, R. AND SOULIER, P. (2020). *Heavy-tailed Time Series*. Springer, New York.
[22] KUNEGIS, J. (2013). KONECT – The Koblenz Network Collection. In Proc. Int. Conf. on World Wide Web Companion. pp. 1343–1350.
[23] LINDSKOG, F., RESNICK, S. I. AND ROY, J. (2014). Regularly varying measures on metric spaces: Hidden regular variation and hidden jumps. *Probability Surveys* **11**, 270–314.
[24] MÓRI, T. F. (2005). The maximum degree of the Barabási–Albert random tree. *Combinatorics, Probability and Computing* **14**, 339–348.
[25] MÓRI, T. F. (2007). Degree distribution nearby the origin of a preferential attachment graph. *Electronic Communications in Probability* **12**, 276–282.
[26] PEKÖZ, E. A., RÖLLIN, A. AND ROSS, N. (2016). Generalized gamma approximation with rates for urns, walks and trees. *The Annals of Probability* **44**, 1776–1816.
[27] PEKÖZ, E. A., RÖLLIN, A. AND ROSS, N. (2017). Joint degree distributions of preferential attachment random graphs. *Advances in Applied Probability* **49**, 368–387.
[28] PEMANTLE, R. (2007). A survey of random processes with reinforcement. *Probability Surveys* **4**, 1–79.
[29] PRATT, J. W. (1960). On interchanging limits and integrals. *The Annals of Mathematical Statistics* **31**, 74–77.
[30] RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer Science & Business Media, New YorkY.
[31] RESNICK, S. I. AND SAMORODNITSKY, G. (2016). Asymptotic normality of degree counts in a preferential attachment model. *Advances in Applied Probability* **48**, 283–299.
[32] TILLIER, C. AND WINTENBERGER, O. (2018). Regular variation of a random length sequence of random variables and application to risk assessment. *Extremes* **21**, 27–56.
[33] VAN DER HOFSTAD, R. (2016). *Random Graphs and Complex Networks*. Cambridge University Press, Cambridge.
[34] VAN DER HOFSTAD, R. (2024). *Random Graphs and Complex Networks*, Vol. **2**. Cambridge University Press, Cambridge.
[35] VOITALOV, I., VAN DER HOORN, P., VAN DER HOFSTAD, R. AND KRIOUKOV, D. (2019). Scale-free networks well done. *Phys. Rev. Res.* **1**, 033034.
[36] WANG, T. AND RESNICK, S. I. (2018). Multivariate regular variation of discrete mass functions with applications to preferential attachment networks. *Methodology and Computing in Applied Probability* **20**, 1029–1042.
[37] WANG, T. AND RESNICK, S. I. (2022). Asymptotic dependence of in-and out-degrees in a preferential attachment model with reciprocity. *Extremes* **25**, 417–450.

https://doi.org/10.1017/apr.2024.73 Published online by Cambridge University Press