The Annals of Applied Probability
2015, Vol. 25, No. 5, 2462–2502
DOI: 10.1214/14-AAP1053

TWITTER EVENT NETWORKS AND THE SUPERSTAR MODEL

BY SHANKAR BHAMIDI[1], J. MICHAEL STEELE AND TAUHID ZAMAN

University of North Carolina, University of Pennsylvania and Massachusetts Institute of Technology

Condensation phenomenon is often observed in social networks such as Twitter where one "superstar" vertex gains a positive fraction of the edges, while the remaining empirical degree distribution still exhibits a power law tail. We formulate a mathematically tractable model for this phenomenon that provides a better fit to empirical data than the standard preferential attachment model across an array of networks observed in Twitter. Using embeddings in an equivalent continuous time version of the process, and adapting techniques from the stable age-distribution theory of branching processes, we prove limit results for the proportion of edges that condense around the superstar, the degree distribution of the remaining vertices, maximal nonsuperstar degree asymptotics and height of these random trees in the large network limit.

1. Retweet graphs and a mathematically tractable model.

Our goal here is to provide a simple model that captures the most salient features of a natural graph that is determined by the Twitter traffic generated by public events. In the Twitter world (or Twitterverse), each user has a set of followers; these are people who have signed-up to receive the tweets of the user. Here, our focus is on *retweets*; these are tweets by a user who forwards a tweet that was received from another user. A retweet is sometimes accompanied with comments by the retweeter.

Let us first start with an empirical example that contains all the characteristics observed in a wide array of such retweet networks. Data was collected during the Black Entertainment Television (BET) Awards of 2010. We first considered all tweets in the Twitterverse that were posted between 10 AM and 4 PM (GMT) on the day of the ceremony, and we then restricted attention to all the tweets in the Twitterverse that contained the term "BET Awards." We view the posters of these tweets as the vertices of an undirected simple graph where there is an edge between vertices v and w if w retweets a tweet received from v, or vice-versa. We call this graph the *retweet graph.*

In the retweet graph for the 2010 BET Awards, one finds a single giant component (see Figure 1). There are also many small components (with five or fewer vertices) and a large number of isolated vertices. The giant component is also ap-

Received November 2012; revised July 2014.

[1]Supported in part by NSF-DMS Grants 1105581 and 1310002 and SES-1357622.

MSC2010 subject classifications. 60C05, 05C80, 90B15.

Key words and phrases. Dynamic networks, preferential attachment, continuous time branching processes, characteristics of branching processes, multitype branching processes, Twitter, social networks, retweet graph.

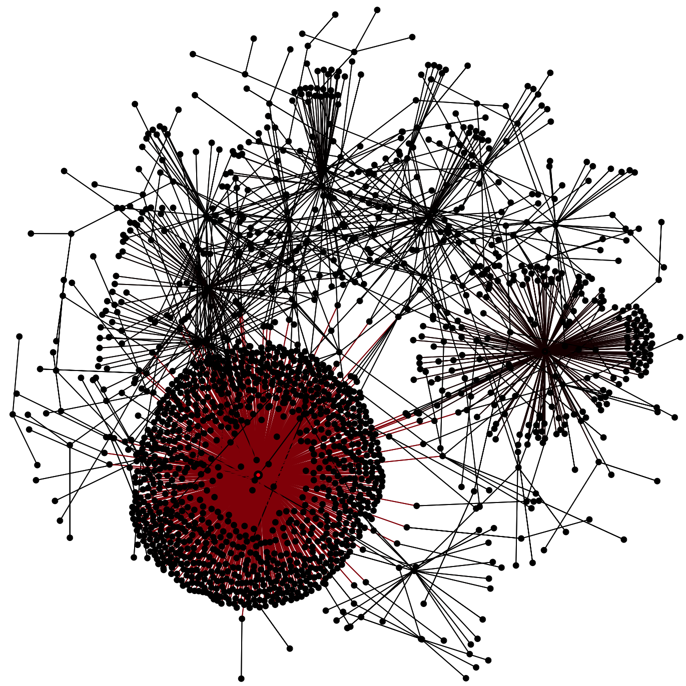

FIG. 1. *Giant component of the* 2010 *BET Awards retweet graph.*

proximately a tree in the sense that if we remove 91 edges from the graph of 1724 vertices and 1814 edges we obtain an honest tree. Finally, the most compelling feature of this empirical tree is that it has one vertex of *exceptionally* large degree. This "superstar" vertex has degree 992, so it is connected to more than 57% of the vertices. As it happens, this "superstar" vertex corresponds to the pop-celebrity Lady Gaga who received an award at the ceremony.

1.1. *Superstar model for the giant component.* Our main observation is that the qualitative and quantitative features of the giant component in a wide array of retweet graphs may be captured rather well by a simple one-parameter model. The construction of the model only makes an obvious modification of the now classic preferential attachment model, but this modification turns out to have richer consequences than its simplicity would suggest. Naturally, the model has the "superstar" property baked into the cake, but a surprising consequence is that the distribution of the degrees of the nonsuperstar vertices is quite different from what one finds in the preferential attachment model.

Our model is a graph evolution process that we denote by $\{G_n, n = 1, 2, \ldots\}$. The graph G_1 consists of the single vertex v_0, that we call the superstar. The graph G_2 then consists of the superstar v_0, a nonsuperstar v_1, and an edge between the two vertices. For $n \geq 2$, we construct G_{n+1} from G_n by attaching the vertex v_n

to the superstar v_0 with probability $0 < p < 1$ while with probability $q = 1 - p$ we attach v_n to a nonsuperstar according to the classical *preferential attachment rule*. That is, with probability q the nonsuperstar v_n is attached to one of the nonsuperstars $\{v_1, v_2, \dots, v_{n-1}\}$ with probability that is proportional to the degree of v_i in G_n.

1.2. *Organization of the paper.* In the next section, we state the main results for the Superstar model. In Section 3, we consider previous work on Twitter networks and explore the connection between our model and existing models. In this section, we also describe two variants of the basic Superstar model (linear attachment and uniform attachment) that can be rigorously analyzed using the same mathematical methodology developed in this paper. In Section 4, we study the performance of this model on various real networks constructed from the Twitterverse and we compare our model to the standard preferential attachment model. Section 5 is the heart of the paper. Here, we construct a special two-type continuous time branching process that turns out to be equivalent to the Superstar model and analyze various structural properties of this continuous time model. In Section 5.2, we prove the equivalence between the continuous time model and the Superstar model through a *surgery* operation. In Section 6, we complete the proofs of all the main results.

2. Mathematical results for the Superstar model. Let $\{G_n, n = 1, 2, \dots\}$ denote the graph process that evolves according to the Superstar model with parameter $0 < p < 1$. We shall think about all the processes constructed on a single probability space through the obvious sequential growth mechanism so that one can make almost sure statements. The degree of the vertex v in the graph G is denoted by $\deg(v, G)$. The first result describes asymptotics of the condensation phenomenon around the superstar. The result is an immediate consequence of the definition of the model and the strong law of large numbers. Since it is a defining element of our model, we set the result out as a theorem.

THEOREM 2.1 (Superstar strong law). *With probability one, we have*

$$\lim_{n\to\infty} \frac{1}{n} \deg(v_0, G_n) = p. \tag{2.1}$$

The next result describes the asymptotic degree distribution.

THEOREM 2.2 (Degree distribution strong law). *With probability one, we have*

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{card}\{1 \le j \le n : \deg(v_j, G_n) = k\} = \nu_{\mathrm{SM}}(k, p),$$

where $\nu_{\mathrm{SM}}(\cdot, p)$ is the probability mass function defined on $\{1, 2, \dots\}$ by

$$\nu_{\mathrm{SM}}(k, p) = \frac{2-p}{1-p}(k-1)! \prod_{i=1}^{k} \left(i + \frac{2-p}{1-p}\right)^{-1}.$$

REMARK 2.3. One should note that the above theorem implies that the degree distribution of the nonsuperstar vertices has a power law tail. Specifically,

$$\frac{2-p}{1-p}(k-1)!\prod_{i=1}^{k}\left(i+\frac{2-p}{1-p}\right)^{-1} \sim C_p k^{-\beta} \qquad \text{as } k \to \infty,$$

for the constants

$$\beta = 3 + p/(1-p), \qquad C_p = \left(\frac{2-p}{1-p}\right)^2 \Gamma\left(\frac{2-p}{1-p}\right),$$

where $\Gamma(x)$ is the gamma function. This should be contrasted with the standard preferential attachment model (with no superstar attachment) whose degree distribution scales like k^{-3} as $k \to \infty$. Thus, although one might expect that this variation in the attachment scheme implies that a fraction $1-p$ of the vertices still continue to perform preferential attachment, and thus the degree distribution should still have a power law exponent of 3; in reality, this attachment scheme has a major effect on the degree distribution. One requires a careful analysis of the different time-scales of the associated continuous time branching process to tease out asymptotic properties of the model.

The next theorem concerns the largest degree amongst all the nonsuperstar vertices $\{v_i : 1 \le i \le n\}$. Let

$$\Upsilon_n := \max_{1 \le i \le n} \deg(v_i, G_n).$$

THEOREM 2.4 (Maximal nonsuperstar degree). *Let* $\gamma = (1-p)/(2-p)$. *There exists a random variable* Δ^* *with* $\mathbb{P}(0 < \Delta^* < \infty) = 1$ *such that*

$$\lim_{n\to\infty} \frac{1}{n^\gamma} \Upsilon_n \xrightarrow{\mathbb{P}} \Delta^*.$$

The almost sure linear growth of the degree of the superstar (Theorem 2.1) is endemic to our construction. For standard preferential attachment (with no superstar attachment mechanism), the maximal degree grows like $\Theta_P(n^{1/2})$ (cf. [19]). Thus, the superstar attachment affects the scaling of the maximal degree as well.

Recall that G_n is a tree. View this tree as rooted at the superstar vertex v_0. Write $\mathcal{H}(G_n)$ for the graph distance of the vertex furthest from the root. Thus, $\mathcal{H}(G_n)$ is the height of the random tree G_n. Theorem 2.1 implies that a fraction p of the vertices in the network are directly connected to the superstar. One might wonder if this reflects a general property of the network, namely does $\mathcal{H}(G_n) = O_p(1)$ as $n \to \infty$? The next theorem shows that in fact the height of the tree increases logarithmically in the size of the network. Let $\mathrm{Lam}(\cdot)$ be the Lambert special function (cf. [9]) and recall that $\mathrm{Lam}(1/e) \approx 0.2784$.

THEOREM 2.5 (Logarithmic height scaling). *With probability one*, *we have*

$$\lim_{n\to\infty} \frac{1}{\log n}\mathcal{H}(G_n) = \frac{1-p}{\mathrm{Lam}(1/e)(2-p)}.$$

3. Related results and questions. In this section, we briefly discuss the connections between this model and some of the more standard models in the literature as well as extensions of the results in the paper. We also discuss previous empirical research done on the structure of Twitter networks.

3.1. *Preferential attachment.* This has become one of the standard workhorses in the complex networks community. It is well nigh impossible to compile even a representative list of references; see [27] where it was introduced in the combinatorics community, [4] for bringing this model to the attention of the networks community, [12, 21] for survey level treatments of a wide array of models, [5] for the first rigorous results on the asymptotic degree distribution and [6, 8, 24] and [13] and the references therein for more general models and results. Let us briefly describe the simplest model in this class of models. One starts with two vertices connected by a single edge as in the Superstar model. Then each new vertex joins the system by connecting to a single vertex in the current tree by choosing this extant vertex with probability proportional to its current degree. In this case, one can show [5] that there exists a limiting asymptotic degree distribution, namely with probability one

$$\lim_{n\to\infty} \frac{1}{n}\,\mathrm{card}\{1 \le j \le n : \deg(v_j, G_n) = k\} = \frac{4}{k(k+1)(k+2)}.$$

Thus, the asymptotic degree distribution exhibits a degree exponent of three. The Superstar model changes the degree exponent of the nonsuperstar vertices from three to $(3-2p)/(1-p)$ (see Theorem 2.4). Further, for the preferential attachment model, the maximal degree scales like $n^{1/2}$ [19], while for the Superstar model, the maximal nonsuperstar degree scales like n^{γ} with $\gamma = (1-p)/(2-p)$.

3.2. *Statistical estimation.* We use real data on various Twitter streams to analyze the empirical performance of the Superstar model and compare this with typical preferential attachment models in Section 4. Estimating the parameters from the data raises a host of new interesting statistical questions. See [29] where such questions were first raised and likelihood based schemes were proposed in the context of usual preferential attachment models. Considering how often such models are used to draw quantitative conclusions about real networks, proving consistency of such procedures as well as developing methodology to compare different estimators in the context of models of evolving networks would be of great interest to a number of different fields.

3.3. *Stable age distribution.* The proofs for the degree distribution build heavily on the analysis of the stable age distribution for a single type continuous time branching process in [20]. We extend this analysis to the context of a two-type variant whose evolution mirrors the discrete type model. Using Perron–Frobenius theory, a wide array of structural properties are known about such models (see [16]). The models used in our proof technique are relatively simpler and we can give complete proofs using special properties of the continuous time embeddings, including special martingales that play an integral role in the treatment (see, e.g., Proposition 5.4). There have been a number of recent studies on various preferential attachment models using continuous time branching processes; see, for example, [2, 11, 25]. For the usual preferential attachment model ($p = 0$), [23] uses embeddings in continuous time and results on the first birth time in such branching processes (see [17]) to show that the height satisfies

$$\frac{\mathcal{H}(\mathcal{G}_n)}{\log n} \xrightarrow{\text{a.s.}} \frac{1}{2\,\mathrm{Lam}(1/e)}.$$

Here, we use a similar technique, but we first need to extend [17] to the setting of multitype branching processes.

3.4. *Previous analysis of Twitter networks.* The majority of work analyzing Twitter networks has been empirical in nature. In one of the earliest studies of Twitter networks [18], the authors looked at the degree distribution of the different networks in Twitter, including retweet networks associated with individual topics. Power-laws were observed, but no model was proposed to describe the network evolution. In [1], the link between maximum degree and the range of time for which a topic was popular or "trending" was investigated. Correlations between the degree in retweet graphs and the Twitter follower graph for different users was studied in [7]. These empirical analyses provided many important insights into the structure of networks in Twitter. However, the lack of a model to describe the evolution of these networks is one of the important unanswered questions in this field, and the rigorous analysis of such a model has not yet been considered. Our work here presents one of the first such models that produces predictions that match Twitter data and also provides a rigorous theoretical analysis of the proposed model.

3.5. *Related models.* One of the main aims of this work is to develop mathematical techniques that extend in a straightforward fashion to variants of the Superstar model. We state results for two such models in this section. We will describe how to extend the proofs for the Superstar model to these variants in Section 6.4. We first start with the *superstar linear preferential attachment*. Fix a parameter $a > -1$. The linear preferential attachment model is described as follows: As before new vertices attach to vertex v_0 with probability p. With probability $q := 1 - p$ the new vertex attaches to one of the extant vertices v, with probability proportional to the $d(v) + a$ where $d(v)$ is the present degree of the vertex. As before, by construction the degree of the superstar v_0 scales like $\sim pn$ as $n \to \infty$. The techniques in the paper extend with simple modifications to prove the following.

THEOREM 3.1 (Linear superstar preferential attachment). *Fix $a > -1$ and $p \in (0, 1)$. In the linear Superstar model, one has for all $k \geq 1$, with probability one*

$$\lim_{n\to\infty} \frac{1}{n}\text{card}\{1 \leq j \leq n : \deg(v_j, G_n) = k\}$$
$$= \frac{2-p+a}{1-p} \frac{\prod_{j=1}^{k-1}(j+a)}{\prod_{i=1}^{k}(i + ((2-p)/(1-p))(1+a))}.$$

Further, for $\gamma(a) = (1-p)/(2-p+a)$, there exists a random variable $0 < \Delta^*(a) < \infty$ a.s. such that the largest degree other than the superstar satisfies

$$n^{-\gamma(a)} \max_{\{1\leq i\leq n\}} \deg(v_i) \xrightarrow{\mathbb{P}} \Delta^*(a) \qquad \textit{as } n \to \infty.$$

Similarly, one can show that the height of the linear Superstar model scales like $\kappa(a)\log n$ for a limit constant $0 < \kappa(a) < \infty$.

We next consider the case of the less realistic Superstar model with *uniform attachment*. Here, each new vertex attaches to the superstar v_0 with probability p or to one of the remaining vertices uniformly at random (irrespective of the degree). Although less realistic in the context of social networks, this is the superstar variant of the random recursive tree a model of a growing tree where each new vertex attaches to a uniformly chosen extant vertex. The random recursive tree has been a model of great interest in the combinatorics and computer science community (see the survey [26]). This model differs from the previous models with the limiting degree distribution possessing exponential tails while the maximal degree only growing logarithmically in the size of the network.

THEOREM 3.2 (Superstar uniform attachment). *Let $q := 1 - p$. For the uniform attachment model, one has for all $k \geq 1$ that with probability one*

$$\lim_{n\to\infty} \frac{1}{n} \text{card}\{1 \leq j \leq n : \deg(v_j, G_n) = k\} = \frac{1}{1+q}\left(\frac{q}{1+q}\right)^{k-1},$$

and the maximal nonsuperstar degree satisfies

$$\lim_{n\to\infty} \frac{\max_{1\leq i\leq n} \deg(v_i)}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\log(1+q)/q}.$$

TABLE 1
For each event E, we list the number of vertices $[|V(G_E^0)|]$, number of edges $[|E(G_E^0)|]$ and maximal degree $[d_{\max}(G_E^0)]$ in the giant component G_E^0, along with the Twitter name of the superstar corresponding to the maximal degree

E	$\|V(G_E^0)\|$	$\|E(G_E^0)\|$	$d_{\max}(G_E^0)$	**Superstar**
1	7365	7620	512	warrenellis
2	3995	4176	362	anison
3	2847	2918	566	FIFAWorldCupTM
4	2354	2414	657	taytorswift13
5	1897	1929	256	FIFAcom
6	1724	1814	992	ladygaga
7	1659	2059	56	MMFlint
8	1408	1459	269	FIFAWorldCupTM
9	1025	1045	247	FIFAWorldCupTM
10	1024	1050	229	SkyNewsBreak
11	705	710	113	realmadrid
12	505	521	186	Wimbledon
13	239	247	38	cnnbrk

4. Retweet graphs for different public events. We collected tweets from the Twitter firehose for thirteen different public events, such as sports matches and musical performances [10]. The Twitter firehose is the full feed of all public tweets that is accessed via Twitter's Streaming Application Programming Interface [28]. By using the Twitter firehose, we were able to access all public tweets in the Twitterverse.

For each public event $E \in \{1, 2, \dots, 13\}$, we kept only tweets that have an event specific term and used those tweets to construct the corresponding retweet graph, denoted by G_E. Our analysis focuses on the giant component of the retweet graph, denoted by G_E^0. In Table 1 we present important properties of each retweet graph's giant component including the number of vertices, number of edges, maximal degree, and the Twitter name of the superstar corresponding to the maximal degree. A more detailed description of each event, including the event specific term, can be found in the Appendix.

The sizes of the giant components range from 239 to 7365 vertices. The giant components of the retweet graphs corresponding to these events are not trees, but they are very tree-like in that they have only a few small cycles. In Table 1, one sees that for each giant component, the deletion of a small number of edges will result in an honest tree.

4.1. *Maximal degree.* The maximal degree in the retweet graphs is larger than would be expected under preferential attachment. Write $n = |V(G_E^0)|$ for the number of vertices in the giant component. For a preferential attachment graph with n

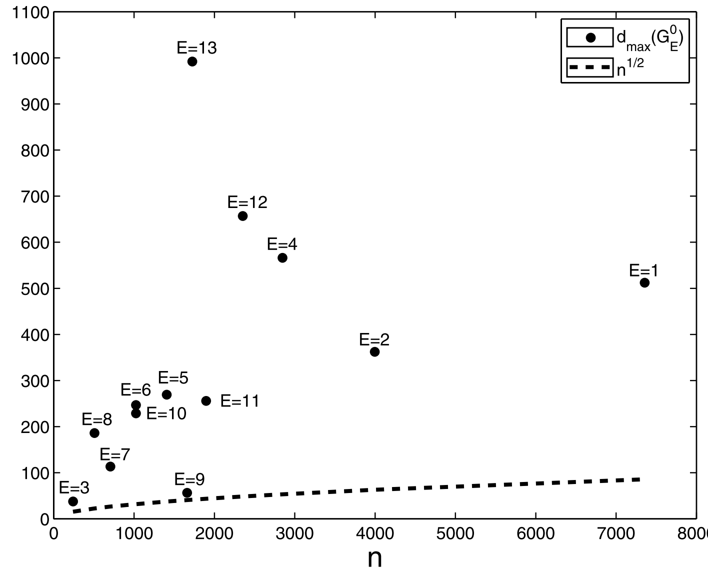

FIG. 2. *Plot of $d_{\max}(G_E^0)$ versus $n = |V(G_E^0)|$ for the retweet graphs of each event. The events are labeled with the same numbers as in Table 1. Also shown is a plot of $n^{1/2}$.*

vertices, it is known that the maximal degree scales as $n^{1/2}$. Figure 2 shows a plot of the maximal degree in the giant component $d_{\max}(G_E^0)$ and a plot of $n^{1/2}$ versus n for the retweet graphs. It can be seen from the figure that the sublinear growth predicted by preferential attachment does not capture the superstar effect in these retweet graphs.

4.2. *Estimating p and the degree distribution.* The asymptotic degree distribution of the Superstar model is known (via Theorem 2.2) once the superstar parameter p is specified. We were interested in seeing, for each event E, how well this model predicted the observed degree distribution in G_E^0. For an event E and degree $k \in \{1, 2, \ldots\}$, we define the empirical degree distribution of the giant component as

$$\widehat{\nu}_E(k) = \frac{1}{|V(G_E^0)|} \operatorname{card}\{v_j \in V(G_E^0) : \deg(v_j, G_E^0) = k\}.$$

To predict the degree distribution using the Superstar model, we need a value for p. We estimate p for each event E as $\widehat{p}(E) = d_{\max}(G_E^0)/|V(G_E^0)|$. Using $p = \widehat{p}(E)$ we obtain the Superstar model degree distribution prediction for each event E and degree k, $\nu_{\mathrm{SM}}(k, \widehat{p})$ from Theorem 2.2. For comparison, we also compare $\widehat{\nu}_E(k)$ to the preferential attachment degree distribution $\nu_{\mathrm{PA}}(k) = 4(k(k+1)(k+2))^{-1}$ [5]. Figure 3 shows the empirical degree distribution for the retweet graphs of four of the events, along with the predictions for the two models. As can be seen, the

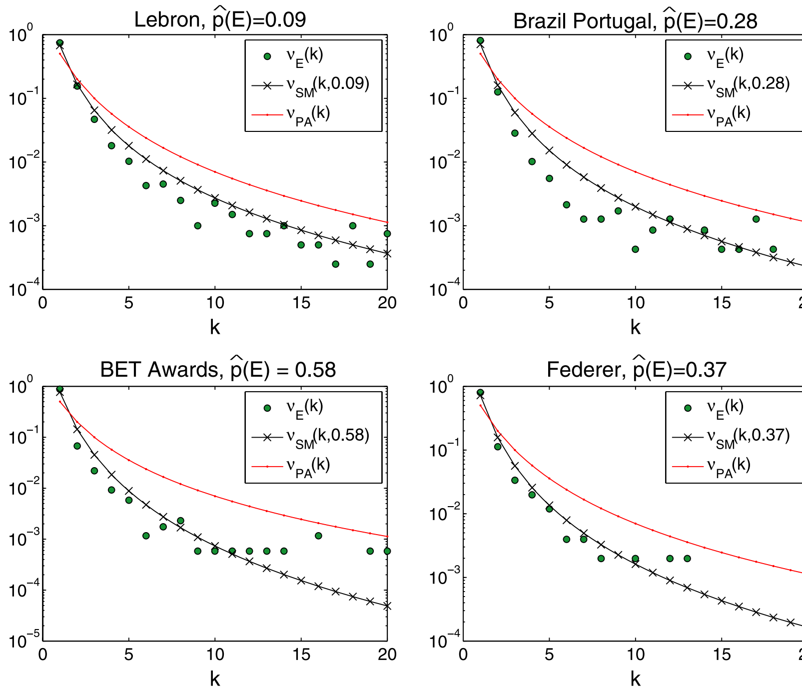

FIG. 3. *Plots of the empirical degree distribution for the giant component of the retweet graphs* [$\nu_E(k)$], *and the estimates of the Superstar model* [$\nu_{\mathrm{SM}}(k, \widehat{p}(E))$] *and preferential attachment* [$\nu_{\mathrm{PA}}(k)$] *for four different events. Each plot is labeled with the event specific term and* $\widehat{p}(E)$.

Superstar model predictions seem to qualitatively match the empirical degree distribution better than preferential attachment. To obtain a more quantitative comparison of the degree distribution, we calculate the relative error of these models for each value of degree k. The relative error for event E and degree k is defined as $\mathrm{relerror}_{\mathrm{SM}}(k, E) = |\nu_{\mathrm{SM}}(k, \widehat{p}) - \widehat{\nu}_E(k)|(\nu_{\mathrm{SM}}(k, \widehat{p}))^{-1}$ for the Superstar model and $\mathrm{relerror}_{\mathrm{PA}}(k, E) = |\nu_{\mathrm{PA}}(k) - \widehat{\nu}_E(k)|(\nu_{\mathrm{PA}}(k))^{-1}$ for preferential attachment. In Figure 4, we show the relative errors for different values of k. As can be seen, the relative error of the Superstar model is lower than preferential attachment for degrees $k = 1, 2, 3, 4$ and for all of the events with the exception of $k = 4$ and $E = 7$. There is a clear connection between the superstar degree and the degree distribution in the giant component of these retweet graphs that is captured well by the Superstar model.

5. Analysis of a special two-type branching process. Let us now start the proofs of the main theorems of Section 2. The core of the proof is a special two-type continuous time branching processes together with a surgery operation that establishes the equivalence between this continuous time construction and the orig-

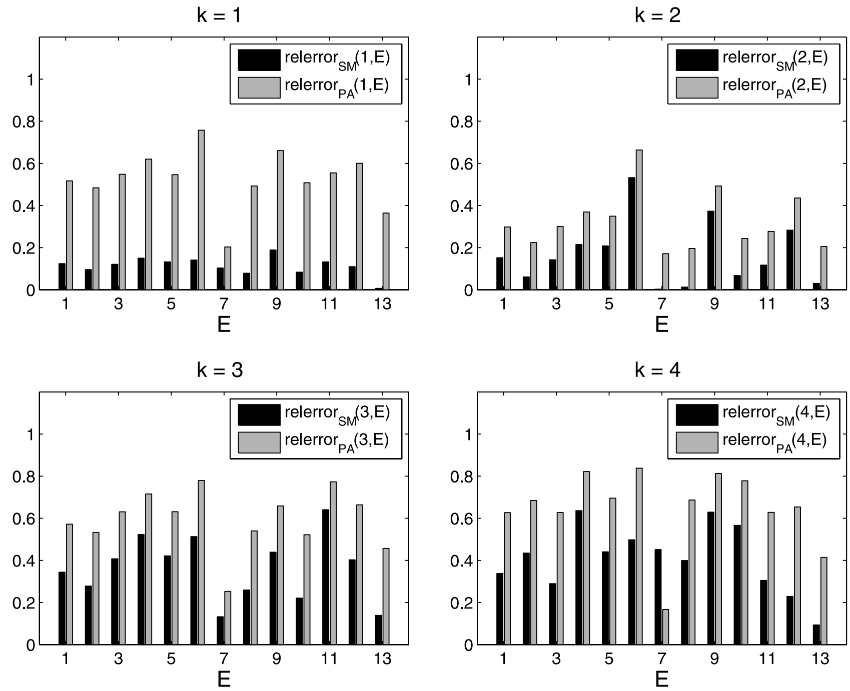

FIG. 4. *Plots of the relative errors of the degree distribution predictions of preferential attachment and the Superstar model for* 13 *retweet graphs. The errors are plotted for degree* $k = 1, 2, 3, 4$.

inal Superstar model. We start by describing this construction and then prove the equivalence between the two models.

5.1. *A two-type continuous branching process.* We now consider a two-type continuous time branching process $\mathsf{BP}(t)$ whose types we call red and blue. For each fixed $t \geq 0$, we shall view $\mathsf{BP}(t)$ as a random tree representing the genealogical structure of the population till time t. This includes parent child relationships of vertices as well as the color of each vertex. We use $|\mathsf{BP}(t)|$ for the total number of individuals in the population by time t. Every individual survives forever. We shall also let $\{\mathsf{BP}(t)\}_{t \geq 0}$ be the associated filtration of the process. Let us now describe the construction. At time $t = 0$, we begin with a single red vertex that we call v_1. For any fixed time $0 < t < \infty$, let V_t denote the vertex set of $\mathsf{BP}(t)$. Each vertex $v \in V_t$ in the branching process gives birth according to a Poisson process with rate

$$\lambda(v, t) = c_B(v, t) + 1,$$

where $c_B(v, t)$ is equal to the number of blue children of vertex v at time t. Also let $c_R(v, t)$ denote the number of red children of vertex v by time t. At the moment of a new birth, this new vertex is colored red with probability p and colored blue with probability $q = 1 - p$. Finally, for $n \geq 1$, define the stopping times

$$\tau_n = \inf\{t : |\mathsf{BP}(t)| = n\}. \tag{5.1}$$

Since the counting process $|\mathsf{BP}(t)|$ is a nonhomogenous Poisson process with a rate that is always greater than or equal to one, the stopping times τ_n are almost surely finite. This completes the construction of the branching process.

5.2. *Equivalence between the branching process and the Superstar model.* Before diving into properties of our two-type branching process constructed as above, let us show how the Superstar model can be obtained from the above branching process via a *surgery* operation. We start with an informal description of the connection between the Superstar model and the branching process $\mathsf{BP}(\cdot)$. To describe this connection, we introduce a new vertex v_0 namely the superstar vertex to the system. Recall that v_1 was the root (the initial progenitor) of the branching process $\mathsf{BP}(\cdot)$. We connect vertex v_1, to the superstar v_0 [v_0 played no role in the evolution of $\mathsf{BP}(\cdot)$]. This forms the Superstar model G_2 on 2 vertices. All the red vertices in the process $\mathsf{BP}(\cdot)$ correspond to the neighbors of the superstar v_0. The true degree of a (nonsuperstar) vertex in G_{n+1} is one plus the number of its blue children in $\mathsf{BP}(\tau_n)$, where the additional factor of one comes from the edge connecting this vertex to its parent. By elementary properties of the exponential distribution, the dynamics of $\mathsf{BP}(\cdot)$ imply that each new vertex that is born is red (connected to the superstar v_0) with probability p, else with probability $q = 1 - p$ is blue and connected to one of the remaining extant (nonsuperstar) vertices with probability proportional to the current degree of that vertex, thus increasing the degree of this chosen vertex by one. These dynamics are the same as the Superstar model.

Formally, our surgery will take the tree $\mathsf{BP}(\tau_n)$ and modify it to get an $(n+1)$-vertex tree $\mathcal{S}_n$ that has the same distribution as the Superstar model G_{n+1}. From this, we will be able to read off the probabilistic properties of the superstar tree G_{n+1}.

We label the vertices of $\mathsf{BP}(\tau_n)$ as $\{v_1, v_2, \ldots, v_n\}$ in order of their birth. Now add a new vertex v_0 to this set to give us the vertex set of the tree $\mathcal{S}_n$. One can anticipate that v_0 will be our superstar. Next, we define the edge set for $\mathcal{S}_n$. To do this, we take each red vertex v in $\mathsf{BP}(\tau_n)$, remove the edge connecting v to its parent (if it has one) and then we create a new edge between v and v_0. To complete the construction of $\mathcal{S}_n$, it only remains to ignore the color of the vertices. An illustration of this surgery for $n = 6$ is given in Figure 5.

PROPOSITION 5.1 (Equivalence from surgery operation). *The sequence of trees $\{\mathcal{S}_n : n \geq 1\}$ has the same distribution as the Superstar model $\{G_{n+1} : n \geq 1\}$.*

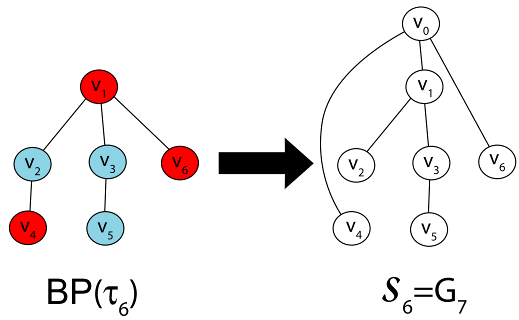

FIG. 5. *The surgery passing from* $\mathsf{BP}(\tau_n)$ *to* $\mathcal{S}_{n+1}$ *and* G_{n+1} *for* $n = 6$.

PROOF. Think of $\mathcal{S}_n$ as being rooted at v_0 so that every vertex except v_0 in $\mathcal{S}_n$ has a unique parent. The parent of all the red individuals is the superstar v_0 while the parents of all of the other blue individuals are unchanged from $\mathsf{BP}(\tau_n)$.

The induction hypothesis will be that $\mathcal{S}_n$ has the same distribution as G_{n+1} and the degree of each nonsuperstar vertex in $\mathcal{S}_n$ is the number of blue children it possesses plus one for the edge connecting the vertex to its parent in $\mathcal{S}_n$. Condition on $\mathsf{BP}(\tau_n)$ and fix $v \in \mathsf{BP}(\tau_n)$. By the property of the exponential distribution, the probability that the next vertex born into the system is born to vertex v is given

$$\frac{\lambda(v, \tau_n)}{\sum_{u \in \mathsf{BP}(\tau_n)} \lambda(u, \tau_n)} = \frac{c_B(v, \tau_n) + 1}{\sum_{u \in \mathsf{BP}(\tau_n)} c_B(u, \tau_n) + 1}.$$

Thus a new vertex v_{n+1} attaches to vertex v with probability proportional to the present degree of v in $\mathcal{S}_n$. Further, with probability p, this vertex is colored red, whence by the surgery operation, the edge to v_{n+1} is deleted and this new vertex is connected to the superstar v_0. In this case the degree of v in $\mathcal{S}_n$ is unchanged. With probability $1 - p$ this new vertex is colored blue, whence the surgery operation does not disturb this vertex so that the degree of vertex v is increased by one. These are exactly the dynamics of G_{n+2} conditional on G_{n+1}. By induction the result follows. □

For the rest of the paper, we shall assume G_{n+1} is constructed through this surgery process from $\mathsf{BP}(\tau_n)$ and suppress $\mathcal{S}_n$.

5.3. *Elementary properties of the branching process.* The previous section set up an equivalence between the Superstar model and the two type continuous time branching process. The aim of this section is to prove properties of this two type branching process. Section 6 uses these results to complete the proof of the main results for the Superstar model.

For $t \geq 0$, write $R(t)$ and $B(t)$ for the total number of red and blue vertices, respectively, in $\mathsf{BP}(t)$. By construction of the process $\{\mathsf{BP}(t) : t \geq 0\}$, every new vertex is independently colored red with probability p and blue with probability $1-p$. In particular, the number of blue vertices $B(t)$ is just a time change of a random walk with Bernoulli($1-p$) increments. Thus, by the strong law of large numbers

$$\frac{B(t)}{|\mathsf{BP}(t)|} \xrightarrow{\text{a.s.}} 1-p \qquad \text{as } t \to \infty. \tag{5.2}$$

Before moving onto an analysis of the branching process, we introduce the Yule process.

DEFINITION 5.2 (Rate a Yule process). Fix $a > 0$. A rate a Yule process is defined as a pure birth process $\mathsf{Yu}_a(\cdot)$ that starts with a single individual $\mathsf{Yu}_a(0) = 1$ and with the rate of creating new individuals proportional to the number of present individuals in the population, namely

$$\mathbb{P}\big(\mathsf{Yu}_a(t+dt) - \mathsf{Yu}_a(t) = 1 \big| \{\mathsf{Yu}_a(s) : 0 \le s \le t\}\big) = a\mathsf{Yu}_a(t)\,dt.$$

The Yule process is a well-studied probabilistic object. The next lemma collects some of its standard properties. In particular, part (a) follows from [22], Section 2.5, while (b) follows from [3], Theorem 1, III.7.

LEMMA 5.3 (Yule process). (a) *For each $t > 0$, the random variable $\mathsf{Yu}_a(t)$ has a geometric distribution with parameter e^{-at}, that is,*

$$\mathbb{P}\big(\mathsf{Yu}_a(t) = k\big) = e^{-at}\big(1-e^{-at}\big)^{k-1}, \qquad k \ge 1.$$

(b) *The process $(e^{-at}\mathsf{Yu}_a(t) : 0 \le t < \infty)$ is an $\mathbb{L}^2$ bounded martingale with respect to the natural filtration and $e^{-at}\mathsf{Yu}_a(t) \xrightarrow{a.s.} W'$, where W' has an exponential distribution with mean one.*

Now define the process

$$M(t) = e^{-(2-p)t}\big(\big|\mathsf{BP}(t)\big| + B(t)\big), \qquad t \ge 0.$$

Note that $M(0) = 1$.

PROPOSITION 5.4 [Asymptotics for $\mathsf{BP}(t)$]. *The process $(M(t) : t \ge 0)$ is a positive $\mathbb{L}^2$ bounded martingale with respect to the natural filtration $\{\mathsf{BP}(t) : t \ge 0\}$, and thus converges almost surely and in $\mathbb{L}^2$ to a random variable W^* with $\mathbb{E}(W^*) = 1$. The random variable W^* is positive with probability one. Further, one has*

$$\lim_{t\to\infty} e^{-(2-p)t}\big|\mathsf{BP}(t)\big| = \frac{W^*}{2-p} \qquad \textit{with probability one}. \tag{5.3}$$

PROOF. We write $Z(t) = |\mathsf{BP}(t)|$ and $Y(t) = Z(t) + B(t)$ so that $M(t) = e^{-(2-p)t}Y(t)$ and we let $dM(t) = M(t+dt) - M(t)$. We then have

$$dM(t) = e^{-(2-p)t}\,dY(t) - (2-p)e^{-(2-p)t}Y(t)\,dt. \tag{5.4}$$

The processes $Z(t)$, $B(t)$ are all counting processes. For such processes, we shall use the infinitesimal shorthand $\mathbb{E}(dZ(t)|\mathsf{BP}(t)) = a(t)\,dt$ to denote the fact that $Z(t) - \int_0^t a(s)\,ds$ is a local martingale.

Now the counting process $Z(t) = |\mathsf{BP}(t)|$ evolves by jumps of size one with

$$\mathbb{P}\big(dZ(t) = 1|\mathsf{BP}(t)\big) = \Big(\sum_{v\in\mathsf{BP}(t)} \big(c_B(v,t)+1\big)\Big)\,dt, \tag{5.5}$$

where $c_B(v,t)$ always denotes the number of blue children of vertex v at time t. The number of blue vertices can be written as $B(t) = \sum_{v\in\mathsf{BP}(t)} c_B(v,t)$ since every blue vertex is an offspring of a unique vertex in $\mathsf{BP}(t)$. Using (5.5) results in

$$\mathbb{E}\big(dZ(t)|\mathsf{BP}(t)\big) = \big(Z(t)+B(t)\big)\,dt.$$

Since $B(t) \le Z(t)$, we see that the rate of producing new individuals is bounded by $2|\mathsf{BP}(t)|$. Thus, the process $|\mathsf{BP}(t)|$ can be stochastically bounded by a Yule process with $a = 2$. This implies by Lemma 5.3 that for all $t \ge 0$ we have $\mathbb{E}(|\mathsf{BP}(t)|^2) < \infty$.

Let us now analyze the process $B(t)$. This process increases by one when the new vertex born into $\mathsf{BP}(\cdot)$ is colored blue that happens with probability $1-p$. Thus, we get

$$\mathbb{E}\big(dB(t)|\mathsf{BP}(t)\big) = (1-p)\big(Z(t)+B(t)\big)\,dt.$$

Combining the last two equation gives us

$$\mathbb{E}\big(dY(t)|\mathsf{BP}(t)\big) = (2-p)Y(t)\,dt.$$

Using (5.4) now gives that $\mathbb{E}(dM(t)|\mathsf{BP}(t)) = 0$. This completes the proof that $M(\cdot)$ is a martingale.

Next, we check that $M(\cdot)$ is an $\mathbb{L}^2$ bounded martingale. Since $Y^2(t+dt)$ can take values $(Y(t)+1)^2$ or $(Y(t)+2)^2$ at rate $pY(t)$ and $(1-p)Y(t)$, respectively, we have

$$\mathbb{E}\big(d\big(M^2(t)\big)|\mathsf{BP}(t)\big) = (4-3p)e^{-(2-p)t}M(t)\,dt.$$

Thus, the process $U(t)$ defined by

$$U(t) = M^2(t) - (4-3p)\int_0^t e^{-(2-p)s}M(s)\,ds,$$

is a martingale. Taking expectations and noting that since $M(\cdot)$ is a martingale, with $M(0) = 1$ thus $\mathbb{E}(M(s)) = 1$ for all s, we get

$$\mathbb{E}\big(M^2(t)\big) = 1 + (4-3p)\int_0^t e^{-(2-p)s}\,ds \le 1 + \frac{4-3p}{2-p}.$$

This $\mathbb{L}^2$ boundedness implies that there exists a random variable W^* such that

$$e^{-(2-p)t}\big(|\mathsf{BP}(t)| + B(t)\big) \overset{\text{a.s.},\mathbb{L}^2}{\longrightarrow} W^*.$$

Using (5.2) shows that $e^{-(2-p)t}|\mathsf{BP}(t)| \to W^*/(2-p)$. To ease notation, write

$$W := \frac{W^*}{(2-p)}.$$

To complete the proof of the proposition we need to show that W is strictly positive. First, note that by $\mathbb{L}^2$ convergence, $\mathbb{E}(W^*) = 1$. So in particular $\mathbb{P}(W = 0) = r < 1$. Let $\zeta_1 < \zeta_2 < \cdots$ be the times of birth of children (blue or red) of the root vertex v_1 and write $\mathsf{BP}_i(\cdot)$ for the subtree consisting of the ith child and its descendants. Then

$$e^{-(2-p)t}|\mathsf{BP}(t)| = \sum_{i=1}^{\infty} e^{-(2-p)\zeta_i}\big[e^{-(2-p)(t-\zeta_i)}|\mathsf{BP}_i(t-\zeta_i)|\big]\mathbb{1}\{\zeta_i \le t\} + e^{-(2-p)t}.$$

Thus, as $t \to \infty$, for any fixed $K \ge 1$, we have

$$W \ge_{\text{st}} \sum_{i=1}^{K} e^{-(2-p)\zeta_i} W_i,$$

where $\{W_i\}_{i\ge1}$ are independent and identically distributed with the same distribution as W (independent of $\{\zeta_i\}_{i\ge1}$) and $\ge_{\text{st}}$ denotes stochastic domination. This independence gives us

$$\mathbb{P}(W = 0) \le \mathbb{P}(W_i = 0\ \forall 1 \le i \le K) = r^K.$$

Letting $K \to \infty$ that $\mathbb{P}(W = 0) = 0$. □

Before ending this section, we derive some elementary properties of the offspring of an individual in $\mathsf{BP}(\cdot)$. Let σ_v be the time of birth of vertex v in $\mathsf{BP}(\cdot)$. Recall that $c_B(v, \sigma_v + s)$ and $c_R(v, \sigma_v + s)$ denote the number of blue and red children, respectively, of this vertex s units of time after the birth of v. Since the distribution of the point process representing offspring of each vertex is the same, these random variables have the same distribution irrespective of the choice of the vertex v. Define the process

$$M^*(t) := c_R(v, \sigma_v + t) - p\int_0^t \big(c_B(v, \sigma_v + s) + 1\big)\,ds, \qquad t \ge 0.$$

LEMMA 5.5 (Offspring point process: distributional properties).

(a) *Conditional on* $\mathsf{BP}(\sigma_v)$ *we have*

$$\big(c_B(v, \sigma_v + t) : t \ge 0\big) \overset{d}{=} \big(\mathsf{Yu}_{1-p}(t) - 1 : t \ge 0\big),$$

and thus one has

$$\mathbb{E}\big(c_B(v, \sigma_v + t)\big) = e^{(1-p)t} - 1, \qquad t \ge 0.$$

(b) *The process* $(M^*(t):t \geq 0)$ *is a martingale with respect to the filtration* $\{\mathsf{BP}(\sigma_v + t):t \geq 0\}$ *and one has*

$$\mathbb{E}\big(c_R(v, \sigma_v + t)\big) = \frac{p}{1-p}\big(e^{(1-p)t} - 1\big), \qquad t \geq 0.$$

PROOF. Part (a) is obvious from construction. To prove (b), note that

$$\mathbb{E}\big(dc_R(v, \sigma_v + t)|\mathsf{BP}(t + \sigma_v)\big) = p\big(c_B(v, \sigma_v + t) + 1\big)\,dt,$$

since vertex v creates a new child at rate $c_B(v, \sigma_v + t) + 1$ which is then marked red with probability p. □

5.4. *Convergence for blue children proportions.* The equivalence between $\mathsf{BP}(\cdot)$ and the Superstar model described in Section 5.2 will imply that the number of vertices with degree $k+1$ in G_{n+1} is the same as the number of vertices in $\mathsf{BP}(\tau_n)$ with exactly k blue children. We will need general results on the asymptotics of such counts for the process $\mathsf{BP}(t)$ as $t \to \infty$. Using the equivalence created by the surgery operation, one can then transfer these results to asymptotics for the degree distribution of the original Superstar model. Now recall the random variable W^* obtained as the martingale limit obtained in Proposition 5.4. Define $p_{\geq k}(\infty)$ as

$$p_{\geq k}(\infty) = k! \prod_{i=1}^{k}\left(i + \frac{2-p}{1-p}\right)^{-1}. \tag{5.6}$$

THEOREM 5.6. *Fix* $k \geq 1$ *and let* $Z_{\geq k}(t)$ *denote the number of vertices in* $\mathsf{BP}(t)$ *that have at least k blue children. Then*

$$e^{-(2-p)t} Z_{\geq k}(t) \xrightarrow{a.s.} p_{\geq k}(\infty)\frac{W^*}{2-p}$$

as $t \to \infty$.

PROOF. The proof uses a variant of the reproduction martingale technique developed in [20] and it is framed in two steps:

(a) Proving convergence of expectations of the desired quantities to the expectations of the asserted limits. This is proved in Section 5.4.1.

(b) Bootstrapping this convergence to almost sure convergence using laws of large numbers. This is proved in Section 5.4.2.

We start with some notation required to carry out this program. For a vertex v, write

$$\boldsymbol{\zeta}^v = \big((\xi_i^v, \mathcal{C}_i^v):i \geq 1\big),$$

for the point process representing offspring (times of birth and types) of this vertex v. More precisely here ξ_i^v denotes the time of birth of the ith offspring of vertex v *after* the birth of vertex v into the branching process $\{\mathsf{BP}(t) : t \geq 0\}$ while $\mathcal{C}_i^v$ denotes the color of this child (red or blue). Thus, the ith offspring of vertex v is born into BP at time $\sigma_v + \xi_i^v$. Write $\boldsymbol{\xi}^v = (\xi_i^v : i \geq 1)$ for the process that just keeps track of times of birth of these offspring for vertex v. Note that the point processes $\boldsymbol{\zeta}^v$ and $\boldsymbol{\xi}^v$ have the same distribution across vertices v. We shall use $\boldsymbol{\zeta} := \boldsymbol{\zeta}^{v_1}$ and $\boldsymbol{\xi} := \boldsymbol{\xi}^{v_1}$ to denote a generic point process with the above distributions. We shall view $\boldsymbol{\xi}$ as a counting measure on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$. For $A \in \mathcal{B}(\mathbb{R}_+)$, write $\boldsymbol{\xi}(A)$ for the number of points in the set A. Define the corresponding intensity measure μ by

$$\mu(A) := \mathbb{E}(\boldsymbol{\xi}(A)), \qquad A \in \mathcal{B}(\mathbb{R}_+).$$

We start with a simple lemma that has notable consequences.

LEMMA 5.7 (Renewal measure). *For $\alpha = 2 - p$, we have*

$$\int_0^\infty e^{-\alpha t} \mu(dt) = 1.$$

The measure defined by setting $\mu_\alpha := e^{-\alpha t}\mu(dt)$ is a probability measure and this measure has expectation $\int_0^\infty t \mu_\alpha(dt) = 1$.

PROOF. As in Lemma 5.5, let $c_B(v_1, t)$ and $c_R(v_1, t)$ denote the number of red and blue children, respectively, of vertex v_1 by time t (note that $\sigma_{v_1} = 0$). Then by definition, the intensity measure μ satisfies $\mu([0, t]) = \mathbb{E}(c_R(v_1, t) + c_B(v_1, t))$. Further by Fubini's theorem,

$$\int_0^\infty e^{-\alpha t} \mu(dt) = \alpha \int_0^\infty e^{-\alpha t} \mu[0, t] \, dt.$$

Using the expressions for $\mathbb{E}(c_B(v_1, t))$ and $\mathbb{E}(c_R(v_1, t))$ from Lemma 5.5 completes the proof. The second assertion regarding the expectation follows similarly. $\square$

5.4.1. *Convergence of expectations*. The first step in the proof of Theorem 5.6 is convergence of expectations. This follows using standard renewal theory. We setup notation that allows us to use the linearity of expectations to derive a renewal equation. We start with the definition of a *characteristic* [14, 15] that we use to count the number of vertices in the branching process with some fixed property. For each vertex $v \in \mathsf{BP}(\infty)$, let $\{\phi^v(s) : s \geq 0\}$ be an independent and identically distributed nonnegative stochastic process, with $\phi^v(s)$ measurable with respect to $\{(\xi_i^v, \mathcal{C}_i^v) : \xi_i^v \leq s\}$. Thus, the value of the stochastic process at time s namely $\phi^v(s)$ is determined by the set offspring of vertex v born before the age s of this vertex v.

The value $\phi^v(s)$ is referred to as the score of vertex v at age s [14], Section 6.9. We write $\phi := \phi^{v_1}$ to denote the process corresponding to the root when we would like to refer to a generic such process. Throughout we shall assume that $\phi(\cdot)$ is bounded and nonnegative, namely for some constant $C < \infty$,

$$\phi(s) \geq 0, \qquad \phi(s) < C \qquad \text{for all } s \geq 0.$$

Define

$$Z_\phi(t) = \sum_{v \in \mathsf{BP}(t)} \phi^v(t - \sigma_v), \qquad t \geq 0$$

for the branching process $\mathsf{BP}(\cdot)$ counted according to characteristic ϕ. The main examples of interest are:

(a) *Total size*: $\phi(s) = 1$ for all $s \geq 0$. This results in $Z_\phi(t) = |\mathsf{BP}(t)|$, the total size of the branching process by time t.

(b) *Degree*: $\phi(s) = \mathbb{1}\{k \text{ or more blue children at age } s\}$ gives $Z_\phi(t) = Z_{\geq k}(t)$, the number of vertices in $\mathsf{BP}(t)$ with k or more blue children.

Now fix an *arbitrary* bounded characteristic ϕ. For fixed time $t > 0$, conditioning on the offspring process $\boldsymbol{\zeta} := \boldsymbol{\zeta}^{v_1}$ of vertex v_1, the branching process counted according to this characteristic satisfies the recursion

$$Z_\phi(t) = \phi^{v_1}(t) + \sum_{\xi_i^{v_1} \leq t} Z_\phi^{(i)}\bigl(t - \xi_i^{v_1}\bigr), \tag{5.7}$$

where $Z_\phi^{(i)}(\cdot) \stackrel{d}{=} Z_\phi(\cdot)$ and are independent for $i \geq 1$ and correspond to the contribution of the descendants of the ith child of vertex v_1. Taking expectations and defining the function $m_\phi(\cdot)$ by $m_\phi(t) := \mathbb{E}(Z_\phi(t))$, this function satisfies the renewal equation

$$m_\phi(t) = \mathbb{E}\bigl(\phi(t)\bigr) + \int_0^t m_\phi(t - s)\mu(ds).$$

Define

$$\tilde{m}_\phi(t) := e^{-\alpha t} m_\phi(t), \qquad t \geq 0.$$

Lemma 5.7 and standard renewal theory ([14], Theorem 5.2.8) now imply the next result.

PROPOSITION 5.8. *For arbitrary bounded characteristics, writing $\alpha = (2 - p)$ we have*

$$\lim_{t \to \infty} \tilde{m}_\phi(t) = \int_0^\infty e^{-\alpha s} \mathbb{E}\bigl(\phi(s)\bigr)\, ds := \tilde{m}_\phi(\infty).$$

Applying this to the two examples which count the size of the branching process and number of vertices with at least k blue children, we get the following result.

COROLLARY 5.9. *Taking the two characteristics of interest one gets for* $\phi(t) = 1$

$$e^{-\alpha t}\mathbb{E}(|\mathsf{BP}(t)|) \to \frac{1}{\alpha} \qquad \text{as } t \to \infty$$

and for $\phi(t) = \mathbb{1}\{k \text{ or more blue children at time } t\}$

$$e^{-\alpha t}\mathbb{E}(Z_{\geq k}(t)) \to \frac{p_{\geq k}(\infty)}{\alpha} \qquad \text{as } t \to \infty,$$

with $p_{\geq k}(\infty)$ *as in* (5.6).

PROOF. The first assertion in the corollary is obvious [corresponding to the case $\phi(\cdot) \equiv 1$]. To prove the second assertion regarding the number of blue vertices, observe that the limit constant in Proposition 5.8 can be written as

$$\begin{aligned}&\frac{1}{\alpha}\int_0^\infty \alpha e^{-\alpha s}\mathbb{E}(\mathbb{1}\{\text{root } v_1 \text{ has } k \text{ or more blue children at age } s\})\,ds\\ &\quad= \frac{1}{\alpha}\mathbb{P}(c_B(v_1, T) \geq k),\end{aligned}$$

where T is an exponential random variable with mean α^{-1} that is independent of the counting process of the number blue offspring $c_B(v_1, \cdot)$. Further, by Lemma 5.5(a),

$$c_B(v_1, \cdot) \stackrel{d}{=} \mathsf{Yu}_{1-p}(\cdot) - 1,$$

where $\mathsf{Yu}_{1-p}(\cdot)$ is rate $1 - p$ Yule process. The interarrival times X_i between blue children i and $i + 1$ are independent exponential random variables with mean $(1 - p)^{-1}(i + 1)^{-1}$, independent of T. In particular $\mathbb{P}(c_B(v_1, T) \geq k) = \mathbb{P}(T > \sum_{j=0}^{k-1} X_j)$. Conditioning on the value of $\sum_{j=0}^{k-1} X_j$ and using tail probabilities for the exponential distribution shows that

$$\mathbb{P}\left(T > \sum_{j=0}^{k-1} X_j\right) = \mathbb{E}\left(\exp\left(-\alpha\sum_{j=0}^{k-1} X_j\right)\right) = \prod_{j=0}^{k-1}\mathbb{E}(\exp(-\alpha X_j)).$$

Using the Laplace transform of the exponential distribution, one can check that the last expression equals $p_{\geq k}(\infty)$. □

5.4.2. *Almost sure convergence.* The aim of this section is to strengthen the convergence of expectations to almost sure convergence. A key role is played by a reproduction martingale, a close relative of the martingale used in [20] to analyze single type branching processes as well as in [17] to analyze times of first birth in generations. Let $v_1, v_2, v_3, \ldots$ denote the vertices of $\mathsf{BP}(\cdot)$ listed in the order of their birth times and let σ_{v_i} denote the time at which vertex v_i is born into the branching process $\mathsf{BP}(\cdot)$. Note that $\sigma_{v_1} = 0$. Recall that $\boldsymbol{\xi}^{v_i} = (\xi_1^{v_i}, \xi_2^{v_i}, \ldots)$

denotes the offspring point process of v_i, namely the first offspring of v_i is born at time $\sigma_{v_i} + \xi_1^{v_i}$, the second offspring of v_i is born at time $\sigma_{v_i} + \xi_2^{v_i}$ and so on. To ease notation, we shall write $\boldsymbol{\zeta}^{(i)} := \boldsymbol{\zeta}^{v_i}$ and $\boldsymbol{\xi}^{(i)} := \boldsymbol{\xi}^{v_i}$. Viewing $\boldsymbol{\xi}^{(i)}$ as a random counting measure on $\mathbb{R}_+$ and writing $\alpha = 2 - p$, we have

$$\xi_\alpha^{(i)} := \sum_{j=1}^{\infty} \exp(-\alpha \xi_j^{v_i}) = \int_0^\infty e^{-\alpha t} \boldsymbol{\xi}^{(i)}(dt).$$

For $m \geq 1$, let $\tilde{\mathcal{F}}_m$ be the sigma-algebra generated by vertices $\{v_1, \dots, v_m\}$ and their offspring processes, namely

$$\tilde{\mathcal{F}}_m := \sigma(\{\boldsymbol{\zeta}^{(i)} : 1 \leq i \leq m\}).$$

For $m = 0$, let $\tilde{\mathcal{F}}_0$ be the trivial sigma-field. Now define $\tilde{R}_0 = 1$ and for $m \geq 0$ define

$$\tilde{R}_{m+1} := \tilde{R}_m + e^{-\alpha \sigma_{v_{m+1}}} (\xi_\alpha^{(m+1)} - 1).$$

Let Γ_m be the set of the first m individuals born and *all* of their offspring. One can check that

$$\tilde{R}_m = \sum_{v \in \Gamma_m} e^{-\alpha \sigma_v} - \sum_{j=1}^{m} e^{-\alpha \sigma_{v_j}}. \tag{5.8}$$

Thus, $\tilde{R}_m$ is a weighted sum of children of the first m individuals with weight $e^{-\alpha \sigma_x}$ for vertex x, the individuals $v_1, v_2, \dots, v_m$ being excluded. In particular, $\tilde{R}_m > 0$ for all m. The next lemma shows that the sequence $(\tilde{R}_m : m \geq 0)$ is much more.

PROPOSITION 5.10 (Reproduction martingale). *The sequence* $(\tilde{R}_m : m \geq 0)$ *is a nonnegative* $\mathbb{L}^2$ *bounded martingale with respect to the filtration* $\{\tilde{\mathcal{F}}_m : m \geq 0\}$. *Thus, there exists a random variable* R_∞ *with* $\mathbb{E}(R_\infty) = 1$ *such that* $\tilde{R}_m \to R_\infty$ *almost surely and in* $\mathbb{L}^2$.

PROOF. By the choice of $\alpha = 2 - p$ in Lemma 5.7 for $i \geq 1$, we have $\mathbb{E}(\xi_\alpha^{(i)}) = \int_0^\infty e^{-\alpha t} \mu(dt) = 1$. Further, $\sigma_{v_{m+1}}$ is $\tilde{\mathcal{F}}_m$ measurable while $\xi_\alpha^{(m+1)}$ is independent of $\tilde{\mathcal{F}}_m$. This implies

$$\mathbb{E}(\tilde{R}_{m+1} - \tilde{R}_m | \tilde{\mathcal{F}}_m) = e^{-\alpha \sigma_{v_{m+1}}} \mathbb{E}(\xi_\alpha^{(m+1)} - 1) = 0.$$

By the orthogonality of the increments of the martingale R_m, we see that

$$\mathbb{E}((\tilde{R}_m - 1)^2) \leq \mathbb{E}([\xi_\alpha^{(i)}]^2) \mathbb{E}\left(\sum_{i=1}^{m} e^{-2\alpha \sigma_{v_i}} \right).$$

Thus, to check $\mathbb{L}^2$ boundedness it is enough to check that the right-hand side is bounded. The following lemma bounds the right-hand side of the above equation in two steps and completes the proof. □

LEMMA 5.11. (a) *Let* $\xi_\alpha := \xi_\alpha^{v_1}$ *and assume* $0 < p < 1$. *Then* $\mathbb{E}([\xi_\alpha]^2) < \infty$. (b) *For any* m, $\mathbb{E}(\sum_{i=1}^m e^{-2\alpha\sigma_{v_i}}) \le 1 + \alpha^{-1}$.

PROOF. To prove (a), we observe that $\xi_\alpha = \int_0^\infty \alpha e^{-\alpha t} \boldsymbol{\xi}[0,t]\,dt$ where $\boldsymbol{\xi}$ is the point process encoding times of birth of offspring of v_1. Thus, by Jensen's inequality with the probability measure $\alpha e^{-\alpha t}\,dt$ we have

$$[\xi_\alpha]^2 \le \int_0^\infty \alpha e^{-\alpha t} [\xi[0,t]]^2\,dt.$$

Let T be an exponential random variable with mean α^{-1} independent of ξ. Thus, it is enough to show $\mathbb{E}([\xi[0,T]]^2) < \infty$. Note that $\xi[0,T] = c_R(v_1,T) + c_B(v_1,T)$, that is, the number of red and blue vertices born to v_1 by the random time T. Thus, it is enough to show $\mathbb{E}(c_R^2(v_1,T))$ and $\mathbb{E}(c_B^2(v_1,T)) < \infty$. Conditioning on $T = t$ first note by using Lemma 5.3 that for fixed t, $\mathbb{E}(c_B^2(v_1,t)) \le Ce^{2(1-p)t}$ where $C < \infty$ is a constant independent of t. Further again using Lemma 5.3, for any fixed t, conditional on $c_B(v_1,t)$, $c_R(v_1,t)$ is stochastically dominated by a Poisson random variable with rate $tc_B(v_1,t)$. Noting that $\alpha = 2 - p$, we get

$$\mathbb{E}([\xi[0,T]]^2) \le C' \int_0^\infty e^{-(2-p)t}\left(e^{2(1-p)t} + t^2 e^{2(1-p)t}\right)dt < \infty,$$

for some constant $C' < \infty$. This completes the proof of (a).

To prove (b), let $S(t) = \sum_{v \in \mathsf{BP}(t)} e^{-2\alpha\sigma_v}$. Then $\sum_{i=1}^m e^{-2\alpha\sigma_{v_i}} = S(\tau_m)$. Further, by (5.5) the rate of creation of new vertices at time t is $|\mathsf{BP}(t)| + B(t)$. Thus, one has

$$\mathbb{E}(dS(t)|\mathsf{BP}(t)) = e^{-2\alpha t}\left(|\mathsf{BP}(t)| + B(t)\right)dt.$$

Taking expectations and noting that $e^{-\alpha t}(|\mathsf{BP}(t)| + B(t))$ is a martingale gives

$$\mathbb{E}(S(t)) = 1 + \int_0^t e^{-\alpha s}\,ds.$$

This completes the proof of part (b), and thus completes the proof of the lemma. □

The next theorem completes the proof of Theorem 5.6. Before stating the main result, we define some new constructs which will be used in the proof. For a bounded characteristic ϕ, recall the limit constant $\tilde{m}_\phi(\infty)$ in Proposition 5.8. In the following theorem, a key role will be played by the martingale $(\tilde{R}_m : m \ge 0)$. Recall that this was a martingale with respect to the filtration $\{\tilde{\mathcal{F}}_m : m \ge 0\}$. We shall switch gears and now think about the process in continuous time. Define $I(t)$ as the set of individuals born after time t whose parents were born before time t and note that

$$\tilde{R}_{|\mathsf{BP}(t)|} = \sum_{x \in I(t)} e^{-\alpha\sigma_x}. \tag{5.9}$$

To ease notation, set

$$R_t := \tilde{R}_{|\mathsf{BP}(t)|}, \qquad \mathcal{F}_t := \tilde{\mathcal{F}}_{|\mathsf{BP}(t)|}. \tag{5.10}$$

THEOREM 5.12 (Convergence of characteristics). *For any bounded characteristic that satisfies the recursive decomposition in* (5.7), *one has*

$$e^{-\alpha t} Z_\phi(t) \xrightarrow{a.s.} \tilde{m}_\phi(\infty) R_\infty.$$

Taking $\phi = 1$ *and using Proposition* 5.4 *implies that* $R_\infty = W^*$, *the a.s. limit of the martingale* $(e^{-\alpha t}(|\mathsf{BP}(t)| + B(t)) : t \geq 0)$.

PROOF. First note that Proposition 5.10 implies that $\{R_t : t \geq 0\}$ is an $\mathbb{L}^2$ bounded martingale with respect to the filtration $\{\mathcal{F}_t : t \geq 0\}$ and thus $R_t \xrightarrow{\text{a.s.}} R_\infty$. For a fixed $c > 0$, define $I(t, c)$ as the set of vertices born after time $(t + c)$ whose parents are born before time t and let

$$R_{t,c} := \sum_{x \in I(t,c)} e^{-\alpha\sigma_x}. \tag{5.11}$$

Obviously, $R_{t,c} \leq R_t$. Intuitively, one should expect $R_{t,c}$ to be small for large c. The next lemma makes this intuition precise. Recall the random variable $\xi_\alpha = \int_0^\infty e^{-\alpha t} \boldsymbol{\xi}(dt)$ where $\boldsymbol{\xi} = \boldsymbol{\xi}^{v_1}$ denoted the point process corresponding to births of offspring of vertex v_1. For fixed $c \geq 0$, write $\xi_\alpha(c) := \int_c^\infty e^{-\alpha t} \boldsymbol{\xi}(dt)$. Finally, define

$$U := \sup_{c \geq 0} e^{c/2} \xi_\alpha(c), \qquad A = \mathbb{E}(U), \qquad K(c) = A e^\alpha \frac{e^{-c/2}}{1 - \sqrt{e}}. \tag{5.12}$$

The proof below will show that $A < \infty$. Also note that $K(c) \to 0$ as $c \to \infty$. Finally, recall from the proof of Proposition 5.4 that we defined $\lim_{t\to\infty} \exp(-\alpha t) \times |\mathsf{BP}(t)| = W$. □

THEOREM 5.13. *For any fixed* $c > 1$, *we have*

$$\limsup_{t\to\infty} R_{t,c} \leq K(c) W \qquad a.s.,$$

where $K(c)$ *is as in* (5.12).

PROOF. The proof uses a variant of the proof used in [20]. Let us start by showing that $\mathbb{E}(U) < \infty$. First, note that for any fixed $c \geq 0$,

$$e^{c/2} \xi_\alpha(c) \leq \int_c^\infty e^{t/2} e^{-\alpha t} \boldsymbol{\xi}(dt) \leq \int_0^\infty e^{t/2} e^{-\alpha t} \boldsymbol{\xi}(dt).$$

Thus, it is enough to show that $\mathbb{E}(\int_0^\infty e^{t/2} e^{-\alpha t} \boldsymbol{\xi}(dt)) < \infty$. By Fubini and integration by parts, $\mathbb{E}(\int_0^\infty e^{t/2} e^{-\alpha t} \boldsymbol{\xi}(dt)) = (\alpha - 1/2) \int_0^\infty e^{t/2} e^{-\alpha t} \mu[0, t]\, dt$ where μ is the intensity measure of the point process $\boldsymbol{\xi}$. Using Lemma 5.5 shows that for some constant $C < \infty$, we have

$$\int_0^\infty e^{t/2} e^{-\alpha t} \mu[0,t]\,dt \le C \int_0^\infty e^{t/2} e^{-\alpha t} e^{(1-p)t}\,dt = C \int_0^\infty e^{-t/2}\,dt < \infty,$$

by using $\alpha = 2 - p$. This completes the proof of finiteness.

Now note that by definition for any $c > 1$

$$R_{t,c} = \sum_{i=1}^{\lfloor t \rfloor} \sum_{\substack{v : \sigma_v \in [i-1,i) \\ j : \xi_j^v + \sigma_v > t+c}} \exp\bigl(-\alpha(\xi_j^v + \sigma_v)\bigr) + \sum_{\substack{v : \sigma_v \in [\lfloor t \rfloor, t) \\ j : \xi_j^v + \sigma_v > t+c}} \exp\bigl(-\alpha(\xi_j^v + \sigma_v)\bigr)$$

$$\le \sum_{i=1}^{\lceil t \rceil} \sum_{\substack{v : \sigma_v \in [i-1,i) \\ j : \xi_j^v + \sigma_v > t+c}} \exp\bigl(-\alpha(\xi_j^v + \sigma_v)\bigr). \tag{5.13}$$

Here, as usual, $\lfloor t \rfloor$ is the largest integer $\le t$ and $\lceil t \rceil$ is the smallest integer $\ge t$. Analogous to the definition of $\xi_\alpha(\cdot)$, define for each vertex v, $\xi_\alpha^v(\cdot)$ using the offspring point process $\boldsymbol{\xi}^v$ of v, namely

$$\xi_\alpha^v(t) := \int_t^\infty \exp(-\alpha t) \boldsymbol{\xi}^v(dt) = \sum_{j : \xi_j^v \ge t} \exp\bigl(-\alpha \xi_j^v\bigr).$$

Further analagous to (5.12), for each vertex v define

$$U_v(t) := e^{t/2} \xi_\alpha^v(t), \qquad U_v := \sup_{t \ge 0} e^{t/2} \xi_\alpha^v(t).$$

Note that

$$U_v \stackrel{d}{=} U, \qquad U_v(t) \le_{\mathrm{st}} U, \tag{5.14}$$

where U is as in (5.12) and $\le_{\mathrm{st}}$ represents stochastic domination. Now for a fixed $i \ge 1$ and vertex v with $\sigma_v \in [i-1, i)$,

$$\sum_{j : \xi_j^v > t+c-\sigma_v} e^{-\alpha(\xi_j^v + \sigma_v)} = e^{-\alpha \sigma_v} \xi_\alpha^v(t + c - \sigma_v) \le e^{-\alpha(i-1)} e^{-(t+c-i)/2} U_v(t + c - \sigma_v).$$

Using this in (5.13) gives

$$R_{t,c} \le \sum_{i=1}^{\lceil t \rceil} e^{-\alpha(i-1)} e^{-(t+c-i)/2} \sum_{v : \sigma_v \in [i-1,i)} U_v(t + c - \sigma_v). \tag{5.15}$$

To proceed, we will need the following generalization of the strong law. We paraphrase the following from [20], Proposition 4.1.

PROPOSITION 5.14 (Extension of the strong law). *Let $\{n_i : i \geq 1\}$ be a sequence of integers and let $(U_{ij} : 1 \leq j \leq n_i)$ be a collection of independent random variables for each fixed $i \geq 1$. Suppose that there exists a random variable $U > 0$ with $\mathbb{E}(U) < \infty$ such that*

$$|U_{ij}| \leq_{\mathrm{st}} U, \qquad 1 \leq j \leq n_i. \tag{5.16}$$

Further assume

$$\liminf_{i\to\infty} \frac{n_{i+1}}{n_1 + \cdots + n_i} > 0. \tag{5.17}$$

Then

$$S_i := \frac{\sum_{j=1}^{n_i}(U_{ij} - \mathbb{E}(U_{ij}))}{n_i} \xrightarrow{a.s.} 0 \qquad \textit{as } i \to \infty, \tag{5.18}$$

and in fact for any $\varepsilon > 0$

$$\sum_{i=1}^{\infty} \mathbb{P}\big(|S_i| > \varepsilon\big) < \infty. \tag{5.19}$$

Proceeding with the proof, for any interval $\mathcal{I} \subseteq \mathbb{R}_+$, write $\mathsf{BP}(\mathcal{I})$ for the collection of vertices born in the interval $\mathcal{I}$ so that $\mathsf{BP}(t) \equiv \mathsf{BP}[0, t]$. We will use the above proposition with $n_i = |\mathsf{BP}[i-1, i)|$ and for each fixed i, the collection of random variables $\{U_v(t + c - \sigma_v) : v \in \mathsf{BP}[i-1, i)\}$. This is a little subtle since the above proposition is stated for deterministic sequences but this justified exactly as in the proof of [20], equation (5.29). First, note that $U_v(t + c - \sigma_v) \leq_{\mathrm{st}} U$ for each fixed v. Note that by Proposition 5.4,

$$\frac{n_{i+1}}{n_1 + \cdots + n_i} := \frac{|\mathsf{BP}[i, i+1)|}{\mathsf{BP}[0, i)} \xrightarrow{\text{a.s.}} e^{\alpha} - 1 > 0,$$

as $i \to \infty$, thus (5.17) is satisfied (almost surely). Using Proposition 5.14 in (5.15) [in particular (5.19)] now shows that for any fixed $\varepsilon > 0$

$$\limsup_{t\to\infty} R_{t,c} \leq \limsup_{t\to\infty} \sum_{i=1}^{\lceil t\rceil} e^{-\alpha(i-1)} e^{-(t+c-i)/2}\big(\mathbb{E}(U) + \varepsilon\big)\big|\mathsf{BP}[i-1, i)\big|.$$

Using the fact that $e^{-\alpha i}|\mathsf{BP}[i-1, i)| \leq e^{-\alpha i}|\mathsf{BP}[0, i)| \xrightarrow{\text{a.s.}} W$, simplifying the above bound and recalling that we used $A = \mathbb{E}(U)$, shows that for every fixed $\varepsilon > 0$

$$\limsup_{t\to\infty} R_{t,c} \leq W(A + \varepsilon)e^{\alpha} \frac{e^{-(c-1)/2}}{1 - \sqrt{e}}.$$

Since ε was arbitrary, this completes the proof. □

COMPLETING THE PROOF OF THEOREM 5.12. Recall that we are dealing with bounded characteristics, that is, $\|\phi\|_\infty < C$ for some constant C. Without loss of generality, let $C = 1$. We shall show that there exists a constant κ such that for all $\varepsilon > 0$,

$$\limsup_{t\to\infty}\left|e^{-\alpha t} Z_\phi(t) - \tilde{m}_\phi(\infty) R_\infty\right| \le \varepsilon(W + 2\kappa R_\infty). \tag{5.20}$$

Since this is true for any arbitrary ε, this completes the proof. Fix $\varepsilon > 0$. First, choose c large such that the bound in Theorem 5.13 satisfies $K(c) < \varepsilon$. Next, for fixed $s > 0$, define the truncated characteristic ϕ_s as

$$\phi_s(u) = \begin{cases} \phi(u), & u \le s, \\ 0, & u > s. \end{cases} \tag{5.21}$$

When the branching process is counted by this characteristic, the contribution of all vertices whose age is more than s is zero. One can view this as a characteristic used to count "young" vertices. The limit constant for this characteristic by Proposition 5.8 is

$$\tilde{m}_{\phi_s}(\infty) = \int_0^s e^{-\alpha u} \mathbb{E}(\phi(u))\, du,$$

where ϕ is the original characteristic. Note that $\tilde{m}_{\phi_s}(\infty) \to \tilde{m}_\phi(\infty)$ as the truncation level $s \to \infty$. Further, writing $\phi' = \phi - \phi_s$, we can view ϕ' as the characteristic counting scores for "old" vertices (vertices of age greater than s). With this notation, we have $Z_\phi(u) = Z_{\phi_s}(u) + Z_{\phi'}(u)$.

Define

$$\tilde{m}_{\phi_s}(u) = e^{-\alpha u} \mathbb{E}(Z_{\phi_s}(u)), \qquad u \ge 0.$$

Now choose $s > c$ large enough with $e^{-\alpha s} < \varepsilon$ such that for all $u > s - c$ one has $e^{-\alpha s} < \varepsilon$, $|\tilde{m}_{\phi_s}(\infty) - \tilde{m}_\phi(\infty)| < \varepsilon$, and $|\tilde{m}_{\phi_s}(u) - \tilde{m}_{\phi_s}(\infty)| < \varepsilon$. The constructs s and c shall remain fixed for the rest of the argument.

Let us understand $Z_{\phi_s}(\cdot)$, the branching process counted according to the truncated characteristic. We first observe that since $\phi_s(u) = 0$ when $u > s$, this implies that for any time $t > s$, vertices born before time $t - s$ (old vertices) do not contribute to $Z_{\phi_s}(t)$. Define $I(t - s)$ as the collection of individuals born after time $t - s$ whose parents were born before time t. Then $Z_{\phi_s}(t)$ decomposes as

$$Z_{\phi_s}(t) = \sum_{v\in I(t-s)} Z^v_{\phi_s}(t - \sigma_v),$$

where $Z^v_{\phi_s}(t - \sigma_v)$ are the contributions to $Z_{\phi_s}(t)$ by the descendants of a vertex v born in the interval $[t - s, t]$. Note that by construction, the parent of such a vertex v belongs to $\mathsf{BP}(t - s)$. Further, recall that in the definition of $R_{t,c}$ in (5.11) we used $I(t - s, c)$ for the set of vertices born after time $(t - s + c)$ whose parents are born before time $t - s$. Then we can further decompose the above sum as

$$Z_{\phi_s}(t) = \sum_{x\in I(t-s)\setminus I(t-s,c)} Z^x_{\phi_s}(t - \sigma_x) + \sum_{x\in I(t-s,c)} Z^x_{\phi_s}(t - \sigma_x).$$

To simplify notation, write $\mathcal{N}(t-s,c) = I(t-s) \setminus I(t-s,c)$, that is, the set of individuals born in the interval $[t-s, t-s+c]$ to parents who were born before time $t-s$. Then we can decompose the difference as a telescoping sum:

$$e^{-\alpha t} Z_\phi(t) - \tilde{m}_\phi(\infty) R_\infty := \sum_{j=1}^{7} E_j(t). \tag{5.22}$$

The definition of these seven terms $\{E_i(t) : 1 \le i \le 7\}$ are as follows:

(a) $E_1(t)$ is defined by setting

$$E_1(t) = e^{-\alpha t} Z_{\phi'}(t), \qquad t \ge 0.$$

Observe that for $E_1(t)$, the only vertices that contribute are those with age greater than s (since $\phi'(u) = 0$ for $u < s$). In particular, $E_1(t) = e^{-\alpha t} Z_{\phi'}(t) \le e^{-\alpha t} |\mathsf{BP}(t-s)|$. Thus, by Proposition 5.4, one has $\limsup_{t\to\infty} E_1(t) \le e^{-\alpha s} W \le \varepsilon W$ a.s. by choice of s.

(b) $E_2(t)$ is defined by setting

$$E_2(t) := \sum_{x \in \mathcal{N}(t-s,c)} e^{-\alpha \sigma_x} \big[e^{-\alpha(t-\sigma_x)} Z^x_{\phi_s}(t-\sigma_x) - \tilde{m}_{\phi_s}(t-\sigma_x) \big].$$

Note that since in the above sum $x \in \mathcal{N}(t-s,c)$, thus $\sigma_x > t-s$. Thus,

$$\big|E_2(t)\big| \le e^{-\alpha(t-s)} |\mathcal{N}(t-s,c)| \frac{\sum_{x \in \mathcal{N}(t-s,c)} e^{-\alpha(t-\sigma_x)} Z^x_{\phi_s}(t-\sigma_x) - \tilde{m}_{\phi_s}(t-\sigma_x)}{|\mathcal{N}(t-s,c)|}.$$

For $E_2(t)$, $\mathcal{N}(t-s,c)$ consists of all children of parents in $\mathsf{BP}(t-s)$ that are born in the interval $[t-s, t-s+c]$. Thus, $|\mathcal{N}(t-s,c)| \le \mathsf{BP}(t-s+c)$. In particular, $\limsup_{t\to\infty} e^{-\alpha(t-s)} |\mathcal{N}(t-s,c)| \le W e^{\alpha c}$. Further, each of the individuals in $\mathsf{BP}(t-s)$ reproduce at rate at least 1. One can check by the strong law of large numbers that $\liminf_{t\to\infty} |\mathcal{N}(t-s,c)|/|\mathsf{BP}(t-s)| \ge c$ almost surely. Finally, the terms in the summand (conditional on $\mathsf{BP}(t-s)$) are independent random variables and each such term in the sum looks like $X - \mathbb{E}(X)$, where X is stochastically bounded by the random variable $Z_{\phi_s}(c)$. A strong law of large numbers argument shows that $\limsup_{t\to\infty} |E_2(t)| = 0$ a.s.

(c) $E_3(t)$ is defined as

$$E_3(t) := \sum_{x \in \mathcal{N}(t-s,c)} e^{-\alpha \sigma_x} \big(\tilde{m}_{\phi_s}(t-\sigma_x) - \tilde{m}_{\phi_s}(\infty) \big).$$

By the choice of s since $t - \sigma_x \ge s - c$, $|\tilde{m}_{\phi_s}(t-\sigma_x) - \tilde{m}_{\phi_s}(\infty)| \le \varepsilon$. Thus, one has $|E_3(t)| \le \varepsilon R_t$. Letting $t \to \infty$, one gets $\limsup_{t\to\infty} |E_3(t)| \le \varepsilon R_\infty$ a.s.

(d) $E_4(t)$ is defined as

$$E_4(t) := \tilde{m}_{\phi_s}(\infty) \bigg(\sum_{x \in \mathcal{N}(t-s,c)} e^{-\alpha \sigma_x} - R_{t-s} \bigg).$$

For $E_4(t)$, we have $|(\sum_{x\in\mathcal{N}(t-s,c)} e^{-\alpha\sigma_x} - R_{t-s})| = R_{t-s,c}$. Thus,

$$\limsup_{t\to\infty} E_4(t) \le \tilde{m}_{\phi_s}(\infty)K(c)W \le \tilde{m}_\phi(\infty)\varepsilon W,$$

almost surely by Theorem 5.13 for the asymptotics of $R_{t,c}$. Here, we have used $\tilde{m}_{\phi_s}(\infty) \le \tilde{m}_\phi(\infty)$ and that our choice of c guarantees $K(c) < \varepsilon$. To ease notation for the rest of the proof, let κ be a constant chosen such that $\max(\sup_{u,s\ge 0}(\tilde{m}_{\phi_s}(u)), \tilde{m}_\phi(\infty)) < \kappa$. The uniform boundedness of ϕ guarantees that this can be done. By choice, κ is independent of s, u. Thus, the bound for the fourth term simplifies to $\limsup_{t\to\infty} E_4(t) \le \kappa\varepsilon W$.

(e) $E_5(t)$ is defined by setting $E_5(t) := \tilde{m}_{\phi_s}(\infty)(R_{t-s} - R_\infty)$. Since $R_{t-s} \xrightarrow{\text{a.s.}} R_\infty$, $E_5(t) \xrightarrow{\text{a.s.}} 0$.

(f) $E_6(t)$ is defined by setting $E_6(t) := R_\infty(\tilde{m}_{\phi_s}(\infty) - \tilde{m}_\phi(\infty))$. By choice of s, $|E_6(t)| \le \varepsilon R_\infty$.

(g) $E_7(t)$ is defined by setting

$$\begin{aligned} E_7(t) &:= e^{-\alpha t} \sum_{v\in I(t-s,c)} Z^v_{\phi_s}(t-\sigma_v) \\ &= \sum_{v\in I(t-s,c)} e^{-\alpha\sigma_v}\bigl(\exp\bigl(-\alpha(t-\sigma_v)\bigr)Z^v_{\phi_s}(t-\sigma_v) - \tilde{m}_{\phi_s}(t-\sigma_v)\bigr) \\ &\quad + \sum_{v\in I(t-s,c)} \exp(-\alpha t)\tilde{m}_{\phi_s}(t-\sigma_v). \end{aligned} \tag{5.23}$$

Using the strong law of large numbers and arguing as in (b) shows that the first term goes to zero as $t\to\infty$ a.s. Using the constant κ defined in (d) above we get

$$\sum_{v\in I(t-s,c)} \exp(-\alpha t)\tilde{m}_{\phi_s}(t-\sigma_x) \le \kappa \sum_{v\in I(t-s,c)} \exp(-\alpha\sigma_x) = \kappa R_{t-s,c}.$$

Using Theorem 5.13 and the choice of c and letting $t\to\infty$, we get

$$\limsup_{t\to\infty} E_7(t) \le \varepsilon\kappa R_\infty \qquad \text{a.s.}$$

Combining all these bounds, one finally arrives at

$$\limsup_{t\to\infty} \bigl|e^{-\alpha t} Z_\phi(t) - \tilde{m}_\phi(\infty)R_\infty\bigr| \le \varepsilon\bigl(W + 2R_\infty + \kappa(W + R_\infty)\bigr) \qquad \text{a.s.}$$

Since $\varepsilon > 0$ was arbitrary, this completes the proof. □

5.5. *Time of first birth asymptotics.* For a rooted tree with root ρ (here $\rho = v_1$), there is a natural notion of a generation of a vertex v. This is defined as the number of edges on the path between v and ρ. Thus, ρ belongs to generation zero, all the neighbors of ρ belong to generation one, and so forth. The aim of this section is to define a modified notion of generation in $\mathsf{BP}(t)$, owing to the fact that the surgery operation as constructed in Section 5.2 that sets up a method to go from the continuous time model to the discrete time model implies that the object of study are the number of edges to the closest red vertex on the path to the root v_1. For each fixed k, we shall define stopping times $\mathrm{Bir}(k)$ representing the first time an individual in modified generation k is born into the process $\mathsf{BP}(\cdot)$. We study asymptotics of $\mathrm{Bir}(k)$ as $k \to \infty$. In the next section, we use these asymptotics to understand height asymptotics for the Superstar model.

Fix $t > 0$. For each vertex $v \in \mathsf{BP}(t)$, let $r(v)$ denote the first red vertex on the path from v to the original progenitor of the process $\mathsf{BP}(\cdot)$, namely v_1. If v is a red vertex then $r(v) = v$. Let $d(v)$ be the number of edges on the path between v and $r(v)$ so that $d(v) = 0$ if v is a red vertex.

Fix $k \geq 1$. Let $\mathrm{Bir}(k)$ denote the stopping times

$$\mathrm{Bir}(k) = \inf\{t > 0 : \exists v \in \mathsf{BP}(t), d(v) = k\}.$$

In other words, $\mathrm{Bir}(k)$ is the first time that there exists a red vertex in $\mathsf{BP}(t)$ such that the subtree consisting of all blue descendants of this vertex and rooted at this red vertex has an individual in generation k. Here, we use Bir to remind the reader that this is the time of the first *birth* in a particular generation. The next theorem proves asymptotics for these stopping times.

THEOREM 5.15. *Let* $\mathrm{Lam}(\cdot)$ *be the Lambert function* [9]. *We have*

$$\frac{\mathrm{Bir}(k)}{k} \xrightarrow{a.s.} \frac{\mathrm{Lam}(1/e)}{1-p} \qquad \text{as } k \to \infty.$$

PROOF. Given any rooted tree $\mathcal{T}$ and $v \in \mathcal{T}$, we shall let $G(v)$ denote the generation of this vertex in $\mathcal{T}$. Write $\mathsf{BP}_b^{v_1}(\cdot)$ for the subtree consisting of all blue descendants of the original progenitor v_1 and rooted at v_1. In distribution, this is just a single type continuous time branching process where each vertex has the same distribution as the process $\mathsf{Yu}_{1-p}(\cdot) - 1$. Further, let

$$\mathrm{Bir}^*(k) = \inf\{t : \exists v \in \mathsf{BP}_b^{v_1}(t), G(v) = k\}.$$

In words, this is the time of first birth of an individual in generation k for the branching process $\mathsf{BP}_b^{v_1}(\cdot)$. From the definitions of $\mathrm{Bir}(k), \mathrm{Bir}^*(k)$, we have $\mathrm{Bir}(k) \leq \mathrm{Bir}^*(k)$.

Much is know about the time of first birth of a single type supercritical branching process, in particular implies that for $\mathsf{BP}_b^{v_1}(\cdot)$, there exists a limit constant β such that

$$\mathrm{Bir}^*(k)/k \xrightarrow{\text{a.s.}} \beta.$$

Here, β can be derived as follows. Write μ_b for the expected intensity measure of the blue offspring, that is, as in Lemma 5.5

$$\mu_b([0, t]) = \mathbb{E}(c_B[v_1, t]) = e^{(1-p)t} - 1, \qquad t \geq 0.$$

For $\theta > 0$, let

$$\Phi(\theta) := \mathbb{E}\left(\int_0^\infty e^{-\theta t} c_B(v_1, dt)\right), \qquad \theta \in \mathbb{R}.$$

It is easy to check that this is finite only for $\theta > 1 - p$ since

$$\Phi(\theta) = \theta \int_0^\infty e^{-\theta t} \mu_b([0,t])\, dt = \frac{1-p}{\theta - (1-p)}.$$

For $a > 0$, define

$$\Lambda(a) := \inf\{\Phi(\theta)e^{\theta a} : \theta \geq 1 - p\} = (1-p)ae^{(1-p)a+1}. \tag{5.24}$$

Then by [17], Theorem 5, the limit constant β is derived as

$$\beta = \sup\{a > 0 : \Lambda(a) < 1\}. \tag{5.25}$$

From this, it follows that $\beta = \mathrm{Lam}(1/e)/(1-p)$ where $\mathrm{Lam}(\cdot)$ is the Lambert function. Then we have

$$\limsup_{k\to\infty} \frac{\mathrm{Bir}(k)}{k} \leq \lim_{k\to\infty} \frac{\mathrm{Bir}^*(k)}{k} \xrightarrow{\text{a.s.}} \frac{W(1/e)}{1-p}.$$

This gives an upper bound in Theorem 5.15. Lemma 5.16 proves a lower bound and completes the proof.

LEMMA 5.16. *Fix any $\varepsilon > 0$ and let $\beta = \mathrm{Lam}(1/e)/(1-p)$ be the asserted limit constant. Then*

$$\sum_{l=1}^{\infty} \mathbb{P}(\mathrm{Bir}(l) < (1-\varepsilon)\beta l) < \infty.$$

Thus, one has $\liminf_{l\to\infty} \mathrm{Bir}(l)/l \geq \beta$ a.s.

PROOF. For ease of notation, for the rest of this proof we shall write $t_\varepsilon(l) = (1-\varepsilon)\beta l$. In the full process $\mathsf{BP}(\cdot)$, two processes occur simultaneously:

(a) New "roots" (red vertices) are created. Recall that we used $R(\cdot)$ for the counting process for the number of red roots.

(b) The blue descendants of each new root have the same distribution as a single type continuous time branching process with offspring process have the same distribution as the process $\mathsf{Yu}_{1-p}(\cdot) - 1$.

Fix $l \geq 2$ and suppose a new red vertex v was created at some time $\sigma_v < t_\varepsilon(l)$. Let $\mathsf{BP}_b^v(\cdot)$ denote the subtree of blue descendants of v. Let $\mathrm{Bir}^*(v, l) > \sigma_v$ be the time of creation of the first blue vertex in generation l for subtree $\mathsf{BP}_b^v(\cdot)$. Now $\mathrm{Bir}(l) < t_\varepsilon(l)$ if and only if there exists a red vertex v born before $t_\varepsilon(l)$ such that the subtree of blue descendants of this vertex has a vertex in generation l

by this time. For a fixed red vertex $v \in \mathsf{BP}(\cdot)$, write $A_v(l)$ for this event. Since $\mathrm{Bir}^*(v,l) - \sigma_v \stackrel{d}{=} \mathrm{Bir}^*(l)$, conditional on $\mathsf{BP}(\sigma_v)$ one has

$$\mathbb{P}\big(A_v(l)|\mathsf{BP}(\sigma_v)\big) = \mathbb{P}\big(\mathrm{Bir}^*(l) \le t_\varepsilon(l) - \sigma_v\big).$$

Fix $0 < s < (1-\varepsilon)\beta l$. Then for $\theta > 1-p$, Markov's inequality implies

$$\mathbb{P}\big(\mathrm{Bir}^*(l) < (1-\varepsilon)\beta l - s\big) \le e^{\theta((1-\varepsilon)\beta l - s)}\mathbb{E}\big[e^{-\theta\,\mathrm{Bir}^*(l)}\big].$$

One of the main bounds of Kingman ([17], equation (2.5), Theorem 1) is $\mathbb{E}[e^{-\theta\,\mathrm{Bir}^*(l)}] \le (\Phi(\theta))^l$. Thus, we get

$$\mathbb{P}\big(\mathrm{Bir}^*(l) < (1-\varepsilon)\beta l - s\big) \le \big[\Phi(\theta)e^{\theta(1-\varepsilon)\beta}\big]^l e^{-\theta s}. \tag{5.26}$$

By the definition of β,

$$\Lambda_\varepsilon := \Lambda\big(\beta(1-\varepsilon)\big) := \inf\big\{\Phi(\theta)e^{\theta(1-\varepsilon)\beta} : \theta > 1-p\big\} < 1,$$

where Λ is as in (5.24). It is easy to check that the minimizer occurs at

$$\theta_\varepsilon = 1 - p + \frac{1}{(1-\varepsilon)\beta}.$$

The final probability bound we shall use is

$$\mathbb{P}\big(\mathrm{Bir}^*(l) < (1-\varepsilon)\beta l - s\big) \le [\Lambda_\varepsilon]^l e^{-\theta_\varepsilon s}. \tag{5.27}$$

Let N_l^ε be the number of red vertices born before time $t_l(\varepsilon)$ whose trees of blue descendants $\mathsf{BP}_b^v(\cdot)$ have at least one vertex in generation l by time $t_\varepsilon(l)$. Obviously, $\mathbb{P}(\mathrm{Bir}(l) < (1-\varepsilon)\beta l) \le \mathbb{E}(N_l^\varepsilon)$. Conditioning on the times of birth of red vertices, one gets

$$\begin{aligned}\mathbb{E}\big(N_l^\varepsilon\big) &\le \int_0^{t_\varepsilon(l)} [\Lambda_\varepsilon]^l d\mathbb{E}\big(R(s)\big) \qquad \text{using equation (5.27)},\\ &= p[\Lambda_\varepsilon]^l \int_0^{t_\varepsilon(l)} e^{-(\theta_\varepsilon - q)s}\,ds \qquad \text{using Lemma 5.5.}\end{aligned}$$

Simplifying, we get for all $l \ge 2$, $\mathbb{E}(N_l^\varepsilon) \le C[\Lambda_\varepsilon]^l$ for a constant C. Thus,

$$\sum_{l=1}^\infty P\big(\mathrm{Bir}(l) < (1-\varepsilon)\beta l\big) < \infty.$$

□

6. Proofs of the main results. Recall the equivalence created by the surgery operation between the Superstar model and the two-type branching process as established in Section 5.2. We shall use this equivalence and the proven results on $\mathsf{BP}(\cdot)$ in Section 5 to complete the proof of the main results. We record the following fact about the asymptotics for the stopping times τ_n.

LEMMA 6.1 (Stopping time asymptotics). *The stopping times* τ_n *satisfy*

$$\tau_n - \frac{1}{2-p}\log n \xrightarrow{a.s.} -\frac{1}{2-p}\log W.$$

PROOF. Proposition 5.4 proves that $|\mathsf{BP}(t)|e^{-(2-p)t} \xrightarrow{\text{a.s.}} W$. Thus $ne^{-(2-p)\tau_n} \xrightarrow{\text{a.s.}} W$. □

Let us now start by proving the main results. We note that Theorem 2.1 is obvious since the degree of the superstar is given by $R(\tau_n) = \sum_{i=1}^n \mathbb{1}\{v_i \text{ is red}\}$, the total number of red vertices and $(\mathbb{1}\{v_i\} \text{ is red})_{i\geq 1}$ is an i.i.d. sequence with Bernoulli p as the marginal distribution. We now prove the remaining results using the correspondence between the continuous time and discrete time processes.

6.1. *Proof of the degree distribution strong law.* In this section, we shall prove Theorem 2.2. Since G_{n+1} is a connected tree, every vertex has degree at least one. Recall that $c_B(v,t)$ denotes the number of blue children of vertex v by time t. Write $\deg(v, G_{n+1})$ for the degree of a vertex in G_{n+1}. The surgery operation implies that for any nonsuperstar vertex

$$\deg(v, G_{n+1}) = c_B(v, \tau_n) + 1. \tag{6.1}$$

Fixing $k \geq 0$, the number of nonsuperstar vertices with degree exactly $k+1$ is the same as the number of vertices in $\mathsf{BP}(\tau_n)$ that have exactly k blue children. Recall that we used $Z_{\geq k}(t)$ for the number of vertices in $\mathsf{BP}(t)$ that have at least k blue children. Proposition 5.4, showed that the total number of vertices $|\mathsf{BP}(t)|$ satisfies

$$e^{-(2-p)t}\big|\mathsf{BP}(t)\big| \xrightarrow{\text{a.s.}} \frac{W^*}{(2-p)} \qquad \text{as } t \to \infty. \tag{6.2}$$

Theorem 5.6 showed that

$$e^{-(2-p)t} Z_{\geq k}(t) \xrightarrow{\text{a.s.}} k! \prod_{i=1}^{k}\left(i + \frac{2-p}{1-p}\right)^{-1} \frac{W^*}{2-p}.$$

Thus, writing $p_{\geq k}(t) = Z_{\geq k}(t)/\mathsf{BP}(t)$ for the proportion of vertices with degree k, Theorem 5.6 implies one has

$$p_{\geq k}(t) \xrightarrow{\text{a.s.}} k! \prod_{i=1}^{k}\left(i + \frac{2-p}{1-p}\right)^{-1} := p_{\geq k}(\infty) \qquad \text{as } t \to \infty.$$

Now let $k \geq 1$. Writing $N_{\geq k}(n)$ for the number of vertices with degree at least k in G_{n+1}, one has $N_{\geq k}(n)/n \xrightarrow{\text{a.s.}} p_{\geq k-1}(\infty)$ as $n \to \infty$. Thus, the proportion of vertices with degree exactly k converges to $p_{\geq k-1}(\infty) - p_{\geq k}(\infty) = \nu_{\mathrm{SM}}(k)$. This completes the proof.

6.2. *Proof of maximal degree asymptotics.* The aim of this is to prove Theorem 2.4. We wish to analyze the maximal nonsuperstar degree that we wrote as

$$\Upsilon_n = \max\{\deg(v_i, G_{n+1}) : 1 \le i \le n\}.$$

The plan will be as follows: we will first prove the simpler assertion of convergence of the degree of vertex v_k for fixed $k \ge 1$. Then we shall show that given any $\varepsilon > 0$, we can choose K such that for large n, the maximal degree vertex has to be one of the first K vertices $v_1, v_2, \ldots, v_K$ with probability greater than $1 - \varepsilon$. This completes the proof.

Fix $k \ge 1$. Recall from (6.1) that $\deg(v_k, G_{n+1}) = c_B(v_k, \tau_n) + 1$ where $c_B(v_k, t)$ are the number of blue vertices born to vertex k by time t. Recall that $c_B(v_k, t)$ is a Yule process of rate $1 - p$ started at time τ_k (i.e., at the birth of vertex v_k). By Lemma 5.3,

$$\frac{c_B(v_k, t)}{e^{(1-p)(t-\tau_k)}} \xrightarrow{\text{a.s.}} W_k', \tag{6.3}$$

where W_k' is an exponential random variable with mean one. Write $\gamma = (1 - p)/(2 - p)$ and let $\Delta_k = e^{-(1-p)\tau_k} W' W^{-\gamma}$. Using (6.2) and (6.3), we have

$$\begin{aligned} n^{-\gamma} \deg(v_k, G_{n+1}) &= \frac{c_B(v_k, \tau_{n-1}) + 1}{e^{(1-p)(\tau_{n-1}-\tau_k)}} \left(\frac{e^{(2-p)\tau_{n-1}}}{|\mathsf{BP}(\tau_{n-1})| + 1} \right)^{\gamma} e^{-(1-p)\tau_k} \\ &\xrightarrow{\text{a.s.}} W_k' W^{-\gamma} e^{-(1-p)\tau_k} := \Delta_k. \end{aligned}$$

Now let us prove distributional convergence of the properly normalized maximal nonsuperstar degree Υ_n. Fix $L > 0$ and let

$$\tilde{M}_n[0, L] := \max\{\deg(v_k, G_{n+1}) : \tau_k \le L\}. \tag{6.4}$$

In other words, this is the largest degree in G_{n+1} amongst all vertices born before time L in $\mathsf{BP}(\cdot)$. The convergence of the degree of v_k for any $k \ge 1$ implies the next result.

LEMMA 6.2 (Convergence near the root). *Fix any $L > 0$. Then there exists a random variable $\Delta^*[0, L] > 0$ such that*

$$n^{-\gamma} \tilde{M}_n[0, L] \xrightarrow{a.s.} \Delta^*[0, L],$$

where $\gamma = (1 - p)/(2 - p)$.

Now if we can show that with high probability, $\Upsilon_n = \tilde{M}_n[0, L]$ for large finite L as $n \to \infty$, then we are done. This is accomplished via the next lemma. Recall that by asymptotics for the stopping times τ_n in Lemma 6.1, given any $\varepsilon > 0$, we can choose $K_\varepsilon > 0$ such that

$$\limsup_{n \to \infty} \mathbb{P}\left(\left| \tau_n - \frac{1}{2 - p} \log n \right| > K_\varepsilon \right) \le \varepsilon. \tag{6.5}$$

For any $0 < L < t$, let $\mathsf{BP}(L, t]$ denote the set of vertices born in the interval $(L, t]$. Recall that we used v_1 for the original progenitor. For any time t and $v \in \mathsf{BP}(t)$, let $\deg_v(t) = c_B(v, t) + 1$ denote the degree of vertex v in the Superstar model $G_{|\mathsf{BP}(t)|+1}$ obtained through the surgery procedure. For fixed K and L, let $A_n(K, L)$ denote the event that for some time $t \in [(2-p)^{-1} \log n \pm K]$, there exists a vertex v in $\mathsf{BP}(L, t]$ with $\deg_v(t) > \deg_{v_1}(t)$.

LEMMA 6.3 (Maxima occurs near the root). *Given any K and ε, one can choose $L > 0$ such that*

$$\limsup_{n\to\infty} \mathbb{P}(A_n(K, L)) \le \varepsilon.$$

In particular, given any $\varepsilon > 0$, we can choose L such that

$$\limsup_{n\to\infty} \mathbb{P}(\Upsilon_n \ne \tilde{M}_n([0, L])) \le \varepsilon.$$

Deferring the proof of this result note that Lemma 6.2 now coupled with the above lemma now shows that there exists a random variable Δ^* such that $\Upsilon_n/n^\gamma \xrightarrow{\mathbb{P}} \Delta^*$. This completes the proof of Theorem 2.4.

PROOF OF LEMMA 6.3. For ease of notation, write

$$t_n^- = (2-p)^{-1} \log n - K, \qquad t_n^+ = (2-p)^{-1} \log n + K.$$

Since the degree of any vertex is an increasing process it is enough to show that we can choose $L = L(K, \varepsilon)$ such that as $n \to \infty$, the probability that there is some vertex born in the time interval $[L, t_n^+]$ whose degree at time t_n^+ is larger than the degree of the root v_1 at time t_n^- is smaller than ε. Let $M_{[L,t_n^+]}(t_n^+)$ denote the maximal degree by time t_n^+ of all vertices born in the interval $[L, t_n^+]$. Then for any constant $C > 0$

$$\begin{aligned}\mathbb{P}(A_n(K, L)) &\le \mathbb{P}(\{\deg_{v_1}(t_n^-) < Cn^\gamma\} \cup \{M_{[L,t_n^+]}(t_n^+) > Cn^\gamma\}) \\ &\le \mathbb{P}(\deg_{v_1}(t_n^-) < Cn^\gamma) + \mathbb{P}(M_{[L,t_n^+]}(t_n^+) > Cn^\gamma).\end{aligned}$$

Since the offspring process of v_1 has the same distribution as a rate $(1-p)$ Yule process

$$e^{-(1-p)t_n^-} \deg_{v_1}(t_n^-) = e^{(1-p)K/2} \frac{\deg_{v_1}(t_n^-)}{n^\gamma} \xrightarrow{\text{a.s.}} W_{v_1},$$

where W_{v_1} has an exponential distribution with mean one. Thus, for a fixed K, we can choose $C = C(\varepsilon)$ large enough such that

$$\limsup_{n\to\infty} \mathbb{P}(\deg_{v_1}(t_n^-) < Cn^\gamma) \le \varepsilon/2.$$

Thus, for a fixed ε, C, K, it is enough to choose L large such that

$$\limsup_{n\to\infty} \mathbb{P}\big(M_{[L,t_n^+]}(t_n^+) > Cn^\gamma\big) \le \varepsilon/2.$$

Without loss of generality, we shall assume throughout that L_ε and t_n^+ are integers. For any integer $L_\varepsilon < m < t_n^+ - 1$, let $M_{[m,m+1]}(t_n^+)$ denote the maximum degree by time t_n^+ of all vertices born in the interval $[m, m+1]$. Then

$$M_{[L,t_n^+]}(t_n^+) = \max_{L\le m\le t_n^+-1} M_{[m,m+1]}(t_n^+).$$

Let $|\mathsf{BP}[m, m+1]|$ denote the number of vertices born in the time interval $[m, m+1]$. Since for a vertex born at some time $s < t_n^+$, the degree of the vertex at time t_n^+ has distribution $\mathrm{Yu}_{1-p}(t_n^+ - s)$, an application of the union bound yields

$$\mathbb{P}\big(M_{[L,t_n^+]}(t_n^+) > Cn^\gamma\big) \le \sum_{m=L}^{t_n^+-1} \mathbb{E}\big(|\mathsf{BP}[m, m+1]|\big)\mathbb{P}\big(\mathrm{Yu}_{1-p}(t_n^+ - m) > Cn^\gamma\big).$$

Now $\mathbb{E}(\mathsf{BP}[m, m+1]) \le \mathbb{E}(|\mathsf{BP}(m+1)|)$. By Proposition 5.4, $\mathbb{E}(|\mathsf{BP}(t)|) \le e^{(2-p)t}$. Further by Lemma 5.3, for fixed time s, a rate $1-p$ Yule process has a geometric distribution with parameter $e^{-(1-p)s}$. Thus, we have

$$\begin{aligned}\mathbb{P}\big(M_{[L,t_n^+]}(t_n^+) > Cn^\gamma\big) &\le \sum_{m=L}^{t_n^+-1} Ae^{(2-p)m}\big[1 - e^{-(1-p)(t_n^+-m)}\big]^{Cn^\gamma} \\ &\le \sum_{m=L}^{t_n^+-1} Ae^{((2-p)m - Ce^{(1-p)(m-K)})},\end{aligned} \tag{6.6}$$

where last inequality follows from the fact that for $0 \le x \le 1$, $1 - x \le e^{-x}$ and

$$e^{t_n^+/2} = n^\gamma e^{(1-p)K}.$$

Now choosing L large, one can make the right-hand side of the last inequality as small as one desires and this completes the proof. □

6.3. *Proof of logarithmic height scaling.* The aim of this section is to complete the proof of Theorem 2.5. Let us first understand the relationship between the distances in $\mathsf{BP}(\tau_n)$ and G_{n+1} due to the surgery operation. The distance of all the red vertices in $\mathsf{BP}(\tau_n)$ from the superstar v_0 is one. For each blue vertex $v \in \mathsf{BP}(\tau_n)$, let $r(v)$ denote the first red vertex on the path from v to the root v_1 in $\mathsf{BP}(\tau_n)$. Recall from Section 5.5 that $d(v)$ denoted the number of edges on the path between v and $r(v)$ with $d(v) = 0$ if v was a red vertex. Then the distance of this vertex from the superstar v_0 in G_{n+1} is just $d(v) + 1$ since the vertex needs $d(v)$ steps to get to $r(v)$ that is then directly connected to v_0 in G_{n+1} by an edge. Let $D(u, v)$ denote the graph distance between vertices u and v in G_{n+1}. Since by convention $d(v) = 0$ for all the red vertices, this argument shows that for all $v \neq v_0 \in G_{n+1}$, $D(v, v_0) = d(v) + 1$. In particular, the height of G_{n+1} is given by

$$\mathcal{H}(G_{n+1}) = \max\{d(v) + 1 : v \in \mathsf{BP}(\tau_n)\}. \tag{6.7}$$

Now by the definition of $\mathcal{H}(G_{n+1})$, there is a vertex in $\mathsf{BP}(\tau_n)$ such that $d(v) = \mathcal{H}(G_{n+1}) - 1$ but no vertex with $d(v) = \mathcal{H}(G_{n+1})$. Recall the stopping times $\mathrm{Bir}(k)$, defined as the first time a vertex with $d(v) = k$ is born in $\mathsf{BP}(\cdot)$. Thus, we have

$$\mathrm{Bir}(\mathcal{H}(G_{n+1}) - 1) \leq \tau_n \leq \mathrm{Bir}(\mathcal{H}(G_{n+1})). \tag{6.8}$$

Now recall that Theorem 5.15 showed that the stopping times $\mathrm{Bir}(k)$ satisfy

$$\mathrm{Bir}(k)/k \xrightarrow{\text{a.s.}} \mathrm{Lam}(1/e)/(1-p) \qquad \text{as } k \to \infty.$$

Dividing (6.8) throughout by $\mathcal{H}(G_{n+1})$ by Theorem 5.15

$$\frac{\mathrm{Bir}(\mathcal{H}(G_{n+1}) - 1)}{\mathcal{H}(G_{n+1})} \xrightarrow{\text{a.s.}} \frac{\mathrm{Lam}(1/e)}{1-p},$$

while by Lemma 6.1 we get

$$\frac{\tau_n}{\log n} \xrightarrow{\text{a.s.}} \frac{1}{2-p}.$$

Rearranging shows that

$$\frac{\mathcal{H}(G_{n+1})}{\log n} \xrightarrow{\text{a.s.}} \frac{(1-p)}{\mathrm{Lam}(1/e)(2-p)}.$$

This completes the proof. □

6.4. *Extension to the variants of the Superstar model.* We now describe how the above methodology easily extends to the two variants described in Section 3, namely the superstar linear preferential attachment and the uniform attachment model (Theorems 3.1 and 3.2). Since the proofs are identical to the original model, modulo the driving continuous time branching process, we will not give full proofs but rather describe the continuous time versions that need to be analyzed to understand the corresponding discrete model. The surgery operation and the subsequent analysis of the continuous time model are identical to the original Superstar model.

For fixed $a > -1$ and $p \in (0, 1)$, we write $\{G_n^{\mathrm{lin}}(a, p) : n \geq 1\}$ for the corresponding family of growing random trees obtained via following the dynamics of the linear attachment scheme (see Section 3). We let $\{G_n^{\mathrm{uni}}(p) : n \geq 1\}$ be the family of random trees obtained via uniform attachment. Now recall that the analysis of the superstar preferential attachment model start with the formulation of a continuous time two type branching process (consisting of red and blue vertices). One then performs surgery on this two type branching process at appropriate stopping times τ_n as defined in (5.1) to obtain the Superstar model. For the two variants, let us now describe the corresponding continuous time versions.

(a) *Superstar linear preferential attachment*: We write $\{\mathsf{BP}_{\text{lin}}(t)\}_{t\geq 0}$ for this branching process. Here one starts with a single red vertex v_1 at time $t = 0$. Each individual lives forever. For any fixed $t \geq 0$, each individual $v \in \mathsf{BP}_{\text{lin}}(t)$ in the branching process reproduces at rate

$$\lambda(v,t) := c_B(v,t) + 1 + a,$$

where as before $c_B(v,t)$ denotes the number of blue children of vertex v at time t. Each new offspring is colored red with probability p and blue with probability $q := 1 - p$.

(b) *Uniform attachment*: Start with a single red vertex v_1 at time $t = 0$. Each individual reproduces at rate one and lives forever. Each new offspring is colored red with probability p and blue with probability $q := 1 - p$. Write $\{\mathsf{BP}_{\text{uni}}(t)\}_{t\geq 0}$ for this branching process.

Fix $n \geq 1$ and recall the stopping time τ_n from (5.1), namely the time for the branching process to reach size n. From Section 5.2, recall the surgery operation that takes $\mathsf{BP}(\tau_n)$ to a random tree $\mathcal{S}_n$ on $n + 1$ vertices. The following proposition which is the general analog of Proposition 5.1 showing the equivalence of the continuous time models and the discrete time versions. The result is stated for the linear preferential attachment model, the same result is true using the corresponding branching process for the uniform attachment model.

PROPOSITION 6.4. *Fix $a > -1$ and $p \in (0, 1)$. Let $\{\mathsf{BP}_{\text{lin}}(t) : t \geq 0\}$ be the continuous time two type branching process constructed as above for the superstar linear preferential attachment model with parameters a, p. The sequence of trees $\{\mathcal{S}_n : n \geq 1\}$ obtained by performing the surgery operation on $\{\mathsf{BP}_{\text{lin}}(\tau_n) : n \geq 1\}$ has the same distribution as $\{G_{n+1}(a, p) : n \geq 1\}$.*

Now recall that in the proof of the original Superstar model, a major role was played by Proposition 5.4 which showed that the associated continuous time branching process grew at rate $\exp((2 - p)t)$. This allowed us to make rigorous the following two ideas (see, e.g., the proof of Corollary 5.9):

(a) As $t \to \infty$, the age of an individual chosen uniformly at random from the population has an exponential distribution with rate $(2 - p)$.

(b) For vertex v, let $c_B(v, \sigma_v + t)$ be the number of blue children t units after being born and note that $\{c_B(\sigma_v + t) : t \geq 0\}$ has the same distribution for any vertex. Since the number of blue children of a vertex represents the out-degree in the Superstar model after the surgery operation, using (i), the limiting degree distribution should be the same as $1 + c_B(v_1, T)$, where $T \sim \exp((2 - p))$ independent of $\{c_B(v_1, t) : t \geq 0\}$. Here, we use v_1 for convenience since $\sigma_{v_1} = 0$.

The corresponding version of Proposition of 5.4 is the following.

PROPOSITION 6.5. (a) *Fix* $a > -1$ *and* $p \in (0, 1)$. *Then there exists a random variable* $W(a, p) > 0$ *a.s. such that as* $t \to \infty$,

$$\exp\bigl(-(2 - p + a)\bigr)\bigl|\mathsf{BP}_{\mathrm{lin}}(t)\bigr| \xrightarrow{a.s.} W(a, p).$$

(b) *For the uniform attachment model, for any* $p \in (0, 1)$ *as* $t \to \infty$,

$$\exp(-t)\bigl|\mathsf{BP}_{\mathrm{uni}}(t)\bigr| \xrightarrow{a.s.} W,$$

where $W \sim \exp(1)$.

PROOF. We start with part (b). For the uniform attachment model, since every individual lives forever and reproduces at rate one, the process $\{|\mathsf{BP}_{\mathrm{uni}}(t)| : t \geq 0\}$ has the same distribution as a rate one Yule process (see Definition 5.2). Then the result follows from Lemma 5.3.

To prove (a), define the process

$$M(t) := \exp\bigl(-(2 - p + a)\bigr)\bigl(\bigl|\mathsf{BP}_{\mathrm{lin}}(t)\bigr| + B(t)\bigr), \qquad t \geq 0,$$

where as before $B(t)$ denotes the number of blue individuals in the population by time t. Arguing exactly as in the proof of Proposition 5.4, it is easy to check that this process is a martingale. The rest of the proof now follows along the same lines as the proof of Proposition 5.4. □

The proof of Theorems 3.1 and 3.2 now proceed as in the analysis of the original model. For example, to show the convergence of the degree distribution for the uniform attachment model Theorem 3.2, first note that for any vertex v, since this vertex reproduces at rate one and each new offspring is colored red with probability p and blue with probability $q = 1 - p$. Thus, the process counting the number of blue children $\{c_B(v_1, t) : t \geq 0\}$ is a rate q Poisson process. Fix $k \geq 1$ and write $Z_{\geq k}(t)$ for the number of vertices in $\mathsf{BP}_{\mathrm{lin}}(t)$ which have k or more blue offspring by time t. The analogous version of Theorem 5.6 for the uniform attachment model implies that

$$\exp(-t)Z_{\geq k}(t) \xrightarrow{\text{a.s.}} p_{\geq k}(\infty)W,$$

where

$$p_{\geq k}(\infty) = \mathbb{P}\bigl(c_B(v, T) \geq k\bigr),$$

where $T \sim \exp(1)$ independent of $c_B(\cdot)$. Now note that

$$\mathbb{P}\bigl(c_B(v, T) \geq k\bigr) = \mathbb{P}\left(\sum_{i=1}^{k} \xi_i \leq T\right),$$

where $\{\xi_i\}_{i \geq 1}$ is a sequence of independent rate q exponential random variables. Arguing as in the proof of Corollary 5.9, we get

$$\mathbb{P}\left(\sum_{i=1}^{k} \xi_i \leq T\right) = \bigl(\mathbb{E}\bigl(\exp(-\xi_i)\bigr)\bigr)^k = \left(\frac{q}{q+1}\right)^k.$$

For the maximal degree, note that by Proposition 6.5 implies that the stopping time τ_n as in (5.1) for the time the continuous time branching process grows to be of size n satisfies

$$\tau_n = \log n + O_P(1).$$

Since for each vertex, its true degree is the number of blue offspring, as an easy lower bound, the root v_1 by time τ_n should have degree $\sim (1-p)\log n$ (since the process describing the blue offspring of the root is just a rate q Poisson process). To get that $\log n$ is the correct order for the maximal degree and in particular the weak law, one argues as in Section 6.2 [in particular see (6.6)], teasing apart the contribution to this maximal degree of vertices born at various times. The proof of Theorem 3.1 is similar. We omit the details.

APPENDIX

Below we describe each of the thirteen events and show the corresponding event specific term.

- $E = 1$: Brazil vs. Netherlands soccer match from the 2010 World Cup. The term is "Brazil" or "Netherlands."
- $E = 2$: Basketball player Lebron James announcement of signing with the Miami Heat. The term is "Lebron."
- $E = 3$: The 2010 World Cup Kick-Off Celebration Concert. The term is "World Cup."
- $E = 4$: Brazil vs. Portugal soccer match from the 2010 World Cup. The term is "Brazil" or "Portugal."
- $E = 5$: Italy vs Slovakia soccer match from the 2010 World Cup. The term is "Italy" or "Slovakia."
- $E = 6$: The 2010 BET Awards show. The term is "BET Awards."
- $E = 7$: The firing of General Stanly McChrystal by US President Barack Obama. The term is "McChrystal."
- $E = 8$: The 2010 World Cup Opening Ceremony. The term is "World Cup."
- $E = 9$: Mexico vs. South Africa soccer match from the 2010 World Cup. The term is "Mexico."
- $E = 10$: England vs. Slovakia soccer match from the 2010 World Cup. The term is "England."
- $E = 11$: Portugal vs. North Korea soccer match from the 2010 World Cup. The term is "Portugal."
- $E = 12$: Roger Federer's tennis match in the first round of the 2010 Wimbledon tournament. The term is "Federer."
- $E = 13$: The UN imposing sanctions on Iran. The term is "Iran."

Acknowledgements. S. Bhamidi would like to thank the hospitality of the Statistics department at Wharton where this work commenced. We thank two referees for detailed comments which lead to a significant improvement in the readability of the paper.

REFERENCES

[1] ASUR, S., HUBERMAN, B. A., SZABO, G. and WANG, C. (2011). Trends in social media: Persistence and decay. In *AAAI Conference on Weblogs and Social Media*, Barcelona, Spain.

[2] ATHREYA, K. B., GHOSH, A. P. and SETHURAMAN, S. (2008). Growth of preferential attachment random graphs via continuous-time branching processes. *Proc. Indian Acad. Sci. Math. Sci.* **118** 473–494. MR2450248

[3] ATHREYA, K. B. and NEY, P. E. (1972). *Branching Processes*. Springer, New York. MR0373040

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

[5] 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. MR1824277

[6] BOLLOBÁS, B. and RIORDAN, O. M. (2003). Mathematical results on scale-free random graphs. In *Handbook of Graphs and Networks* 1–34. Wiley, Weinheim. MR2016117

[7] CHA, M., HADDADI, H., BENEVENUTO, F. and GUMMADI, K. P. (2010). Measuring user influence in Twitter: The million follower fallacy. In *AAAI Conference on Weblogs and Social Media*, Washington, DC.

[8] COOPER, C. and FRIEZE, A. (2003). A general model of web graphs. *Random Structures Algorithms* **22** 311–335. MR1966545

[9] CORLESS, R. M., GONNET, G. H., HARE, D. E. G., JEFFREY, D. J. and KNUTH, D. E. (1996). On the Lambert W function. *Adv. Comput. Math.* **5** 329–359. MR1414285

[10] DATA COURTESY MICROSOFT RESEARCH CAMBRIDGE. August 2010.

[11] DEIJFEN, M. (2010). Random networks with preferential growth and vertex death. *J. Appl. Probab.* **47** 1150–1163. MR2752883

[12] DOROGOVTSEV, S. N. and MENDES, J. F. F. (2003). *Evolution of Networks*. Oxford Univ. Press, Oxford. MR1993912

[13] DURRETT, R. (2007). *Random Graph Dynamics*. Cambridge Univ. Press, Cambridge. MR2271734

[14] JAGERS, P. (1975). *Branching Processes with Biological Applications*. Wiley, London. MR0488341

[15] JAGERS, P. and NERMAN, O. (1984). The growth and composition of branching populations. *Adv. in Appl. Probab.* **16** 221–259. MR0742953

[16] JAGERS, P. and NERMAN, O. (1996). The asymptotic composition of supercritical multi-type branching populations. In *Séminaire de Probabilités, XXX. Lecture Notes in Math.* **1626** 40–54. Springer, Berlin. MR1459475

[17] KINGMAN, J. F. C. (1975). The first birth problem for an age-dependent branching process. *Ann. Probab.* **3** 790–801. MR0400438

[18] KWAK, H., LEE, C., PARK, H. and MOON, S. (2010). What is Twitter, a social network or a news media? In *Proc. WWW. Proceedings of the* 19*th International Conference on World Wide Web* 591–600. ACM, New York.

[19] MÓRI, T. F. (2007). Degree distribution nearby the origin of a preferential attachment graph. *Electron. Commun. Probab.* **12** 276–282 (electronic). MR2342706

[20] NERMAN, O. (1981). On the convergence of supercritical general (C–M–J) branching processes. *Z. Wahrsch. Verw. Gebiete* **57** 365–395. MR0629532

[21] NEWMAN, M. E. J. (2003). The structure and function of complex networks. *SIAM Rev.* **45** 167–256 (electronic). MR2010377

[22] NORRIS, J. R. (1998). *Markov Chains. Cambridge Series in Statistical and Probabilistic Mathematics* **2**. Cambridge Univ. Press, Cambridge. MR1600720

[23] PITTEL, B. (1994). Note on the heights of random recursive trees and random m-ary search trees. *Random Structures Algorithms* **5** 337–347. MR1262983

[24] RUDAS, A. and TÓTH, B. (2009). Random tree growth with branching processes—a survey. In *Handbook of Large-Scale Random Networks*. *Bolyai Soc*. *Math*. *Stud*. **18** 171–202. Springer, Berlin. MR2582389

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

[26] SMYTHE, R. T. and MAHMOUD, H. M. (1995). A survey of recursive trees. *Theory Probab. Math*. *Statist*. **51** 1–27.

[27] SZYMAŃSKI, J. (1987). On a nonuniform random recursive tree. In *Random Graphs'*85 (*Poznań*, 1985). *North-Holland Math*. *Stud*. **144** 297–306. North-Holland, Amsterdam. MR0930497

[28] TWITTER DEVELOPERS. Available at https://dev.twitter.com/docs/api/1.1/get/statuses/firehose, August 2012.

[29] WIUF, C., BRAMEIER, M., HAGBERG, O. and STUMPF, M. P. H. (2006). A likelihood approach to analysis of network data. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **103** 7566–7570 (electronic). MR2221040

S. BHAMIDI
DEPARTMENT OF STATISTICS
AND OPERATIONS RESEARCH
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599
USA
E-MAIL: bhamidi@email.unc.edu

J. M. STEELE
THE WHARTON SCHOOL
DEPARTMENT OF STATISTICS
HUNTSMAN HALL 447
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104
USA
E-MAIL: steele@wharton.upenn.edu

T. ZAMAN
SLOAN SCHOOL OF MANAGEMENT
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02142
USA
E-MAIL: zlisto@mit.edu