arXiv:1211.3090v1 [math.PR] 13 Nov 2012

TWITTER EVENT NETWORKS AND THE SUPERSTAR MODEL

SHANKAR BHAMIDI AND J.MICHAEL STEELE AND TAUHID ZAMAN

ABSTRACT. Motivated by "condensation" phenomena 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 which 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 non-superstar 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 which contains all the characterstics 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 1.1). There are also many small components (with five or fewer vertices) and a large number of isolated vertices. The giant component is also approximately a tree in

Date: November 14, 2012,
Key words: dynamic networks, preferential attachment, continuous time branching processes, multi-type branching processes, Twitter, social networks, retweet graph.
MSC2000 subject classification. 60C05, 05C80, 90B15.

Shankar Bhamidi: University of North Carolina, Chapel Hill, NC 27599. Email address: `bhamidi@email.unc.edu`.

J. M. Steele: The Wharton School, Department of Statistics, Huntsman Hall 447, University of Pennsylvania, Philadelphia, PA 19104. Email address: `steele@wharton.upenn.edu`.

Tauhid Zaman: Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02142. Email address: `zlisto@mit.edu`.

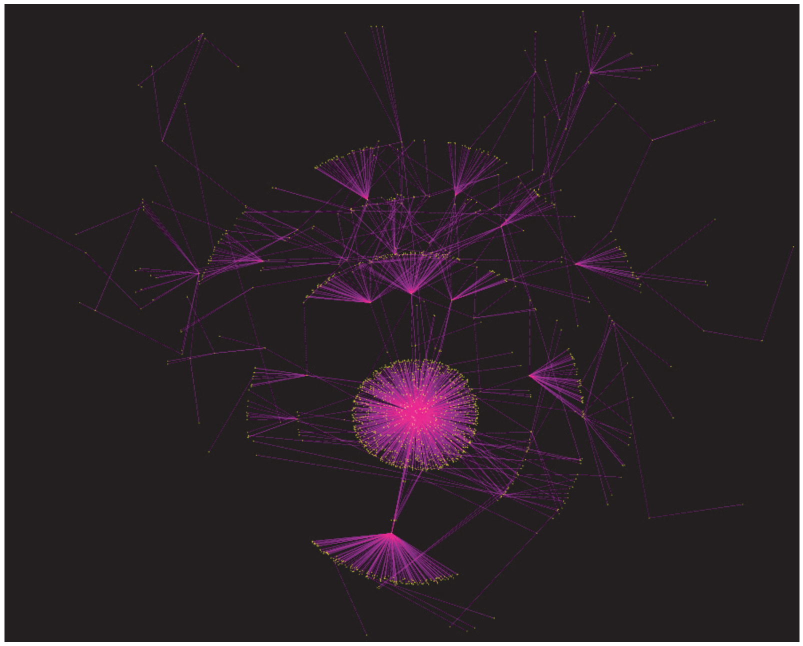

FIGURE 1.1. Giant component of the 2010 BET Awards retweet graph.

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 the giant component of the retweet graph 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 non-superstar vertices is totally different from what one finds in the preferential attachment model.

To construct the model we consider 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, and we call v_0 the superstar. The graph G_2 then consists of the superstar v_0 , a non-superstar v_1, and an edge between the two vertices. For $n \geq 2$, we then construct G_{n+1} from G_n by attaching the vertex v_n to the superstar with probability $0 < p < 1$ while with probability $q = 1-p$ we attach v_n to a non-superstar according to the classical *preferential attachment rule*. That is, with probability q the non-superstar v_n is attached to one of the non-superstars $\{v_1, v_2, \ldots, v_{n-1}\}$, and given that v_n is attached to a non-superstar, it is attached to the vertex v_i, $1 \leq i \leq n-1$, with probability that is proportional to the degree of v_i in G_n.

1.2. **Organization of the paper.** The rest of the paper is organized as follows. In the next section, we state the main mathematical results for the Superstar Model. We discuss previous work analyzing Twitter networks and the connection between the model analyzed in this paper and existing models in Section 3. In Section 4 we study the performance of this model on various real networks constructed from the Twitterverse and compare this to the standard preferential attachment model. Section 5 is the heart of the paper where we construct a special two type continuous time branching process which turns out to be equivalent to the Superstar Model and analyze various structural properties of this continuous time model. In Section 6 we prove the equivalence between the continuous time model and the Superstar Model through a *surgery* operation. In Section 7 we complete the proofs of all the main results by using the equivalence between the two models and the proven properties of the continuous time model to read off results for the Superstar Model.

2. Mathematical Results for the Superstar Model

Let $\{G_n,\ n = 1, 2, \ldots\}$ denote a graph process that follows the 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. As before, the first vertex v_0 is called the "superstar." and the remaining vertices are non-superstars. The degree of the vertex v in the graph G is denoted by $\deg(v, G)$. The first result describes asymptotics of the condensation phenomena around the superstar.

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*

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

where $\nu_{SM}(\cdot, p)$ is the probability mass function defined by

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

Remark 2.3. This theorem implies that the degree distribution of the non-superstar vertices have a power law tail. Specifically,

$$\frac{2-p}{1-p} (k-1)! \prod_{i=1}^{k} \left(i + \frac{2-p}{1-p}\right)^{-1} \sim C_p k^{-\alpha} \quad ,$$

as $k \to \infty$ for the constants $\alpha = (3 - 2p)/(1 - p)$ and $C_p = (2 - p)/(1 - p)\Gamma(\alpha)e^{2+\alpha}$.

The next theorem concerns the largest degree amongst all the non-superstar 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 non-superstar degree).** *Let $\gamma = (1-p)/(2-p)$. There exists a non-degenerate strictly positive finite random variable Δ^* such that with probability one we have*

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

The almost sure linear growth of the degree of the superstar (Theorem 2.1) is to be expected from our construction. The scaling of the second largest degree vertex underscores a notable divergence from the preferential attachment model where the maximal degree grows at the rate $O(n^{1/2})$ [20].

Recall that G_n is a tree. We shall think of this tree as rooted at the superstar v_0. Let $\mathcal{H}(G_n)$ denote the graph distance of the vertex furthest from the root. Call this the height of G_n. Theorem 2.1 implies that a fraction p of the network is directly connected to the superstar. One immediately wonders if this reflects a general property of the network, does the height $\mathcal{H}(G_n) = O_p(1)$ as $n \to \infty$? The next theorem shows that in fact the height of the tree increases logarithmically in the size of the network.

Theorem 2.5 (**Logarithmic height scaling).** *Let $W(\cdot)$ be the Lambert special function with $W(1/e) \approx 0.2784$. Then with probability one we have*

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

3. Related results and questions

The fields of social networks and attachment models have witnessed an explosive growth over the last few years. 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.

(a) **Preferential attachment:** This has become one of the standard workhorses in the complex networks community. It is almost impossible to provide even a partial list of references but see [7] for bringing this model to the attention of the networks community, [22],[13] for survey level treatments of a wide array of models, [9] for the first rigorous results on the asymptotic degree distribution, and [11], [8], [26], and [14] and the references therein for more general models and results. Restricting ourselves to the simplest case, one starts with two vertices connected by a single edge as in the Superstar Model and then each new vertex joins the system by connecting to a single vertex in the current tree by choosing this vertex with probability proportional to its degree. In this case, one can show ([9]) that there exists a limiting asymptotic degree distribution such that with probability one

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

thus exhibiting a degree exponent of three. The Superstar Model changes the degree exponent of the non-superstar vertices from three to $(3-2p)/(1-p)$ (see Theorem 2.4). Further, for the preferential attachment model the maximal degree scales like $n^{1/2}$ ([20]), while for the Superstar Model, the maximal non-superstar degree scales like n^γ with $\gamma = (1-p)/(2-p)$.

(b) **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 [27] 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.

(c) **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 [21]. 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 [17]). 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 which play an integral role in the treatment (see e.g. Proposition 5.3). There have been a number of recent studies on various preferential attachment models using continuous time branching processes, see e.g. [25, 5, 12]. For the usual preferential attachment model ($p = 0$), [24] using embeddings in continuous time and results on the first birth time in such branching processes [18] shows that the height satisfies

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

We use a similar technique but we first need to extend [18] to the multi-type setting, of relevance to us.

(d) **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 [19] 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 [4] 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 [10]. 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 even been considered yet. Our work here presents one of the first such models which produces predictions that match Twitter data and also is given a rigorous theoretical analysis.

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 [1]. The Twitter firehose is the full feed of all public tweets which is accessed via Twitter's Streaming Application Programming Interface [2]. By using the Twitter firehose, we were able to access all public tweets in the Twitterverse.

For each public event $E \in \{1, 2, ..., 13\}$, we kept only tweets which have an event specific term and used those tweets to construct the retweet graph which we denote G_E. Our analysis focuses on the giant component of the retweet graph, which we denote G_E^0. In

E	$\lvert V(G^0_E)\rvert$	$\lvert E(G^0_E)\rvert$	$d_{max}(G^0_E)$	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

TABLE 4.1. For each event E, we list the number of vertices ($|V(G^0_E)|$), number of edges ($|E(G^0_E)|$) , and maximal degree ($d_{max}(G^0_E)$) in the giant component G^0_E, along with the Twitter name of the superstar corresponding to the maximal degree.

Table 4.1 we present important properties of each retweet graph's giant component such as 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 are not trees, but are very tree-like. As can be seen in the table, 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. Let us call the number of vertices in the giant component $n = |V(G^0_E)|$. For a preferential attachment graph with n vertices it is known that the maximal degree scales as $n^{1/2}$. Figure 4.1 shows a plot of the maximal degree in the giant component $d_{max}(G^0_E)$ 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 is not capturing the superstar effect in these retweet graphs.

4.2. **Estimating p and the degree distribution.** The Superstar Model degree distribution is known once the superstar parameter p is specified. We are interested in seeing if for each event E this model can predict the degree distribution in G^0_E. For an event E and degree $k \in \{1, 2, ...\}$ we define the empirical degree distribution of the giant component as

$$\widehat{\nu}_E(k) = \frac{1}{|V(G^0_E)|}\text{card}\left\{v_j \in V(G^0_E) : \deg(v_j, G^0_E) = k\right\}$$

To predict the degree distribution using the Superstar Model, we need a value for p. We estimate p for each event E as $\widehat{p}(E) = d_{max}(G^0_E)/|V(G^0_E)|$. Using $p = \widehat{p}(E)$ we obtain the Superstar Model degree distribution prediction for each event E and degree k, $\nu_{SM}(k, \widehat{p})$ from Theorem 2.2. For comparison, we also compare $\widehat{\nu}_E(k)$ to the preferential attachment

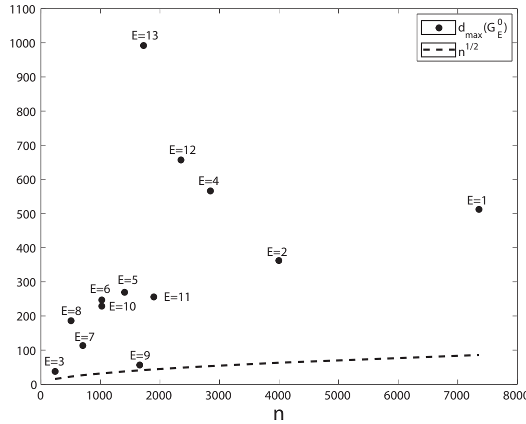

FIGURE 4.1. Plot of $d_{max}(G_E^0)$ versus $n = |V(G_E^0)|$ for the retweet graphs of each event. The events are labeled with the same numbers as in Table 4.1. Also shown is a plot of $n^{1/2}$.

degree distribution $\nu_{PA}(k) = 4\,(k(k+1)(k+2))^{-1}$ [9]. Figure 4.2 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 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 $\text{relerror}_{SM}(k,E) = |\nu_{SM}(k,\widehat{p}) - \widehat{\nu}_E(k)|\,(\nu_{SM}(k,\widehat{p}))^{-1}$ for the Superstar Model and $\text{relerror}_{PA}(k,E) = |\nu_{PA}(k) - \widehat{\nu}_E(k)|\,(\nu_{PA}(k))^{-1}$ for preferential attachment. In Figure 4.3 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 which is captured well by the Superstar Model.

5. Analysis of a special two type branching process

The proofs of the theorems of Section 2 exploit a special two-type continuous time branching processes together with a simple surgery that proves the equivalence between this construction and the superstar model. We start by describing this construction and proving the equivalence between the two models. We shall then derive various properties (degree distribution, height and maximal degree) of the continuous time version and show how these results carry over to the Superstar Model.

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. We use $|\mathsf{BP}(t)|$ for the total number of individuals in the population by time t. In the construction, every individual survives forever so there is no distinction between living and dead individuals. We shall also let $\{\mathsf{BP}(t)\}_{t\geq 0}$ be the associated filtration of the process. At time $t = 0$ we begin with a single red vertex which 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 tree gives birth

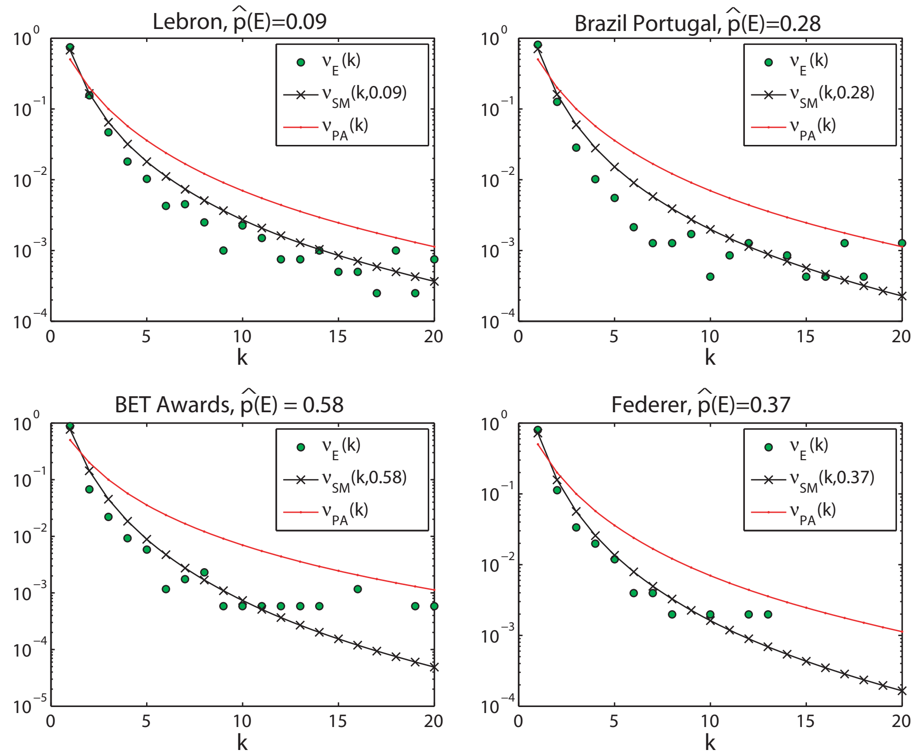

FIGURE 4.2. Plots of the empirical degree distribution for the giant component of the retweet graphs ($\nu_E(k)$), and the estimates of the Superstar Model ($\nu_{SM}(k, \widehat{p}(E))$) and preferential attachment ($\nu_{PA}(k)$) for four different events. Each plot is labeled with the event specific term and $\widehat{p}(E)$.

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 by time t. At the moment of a new birth, the new child vertex is colored red with probability p and colored blue with probability $q = 1-p$. There are no deaths of vertices, and all vertices continue to procreate through all time. For $t \geq 0$, write $R(t)$ and $B(t)$ for the total number of red and blue vertices respectively in $\mathsf{BP}(t)$. Finally for $n \geq 1$, define the stopping times

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

Since the counting process $|\mathsf{BP}(t)|$ is a non-homogenous Poisson process with a rate that is always greater than or equal to one, we see that for any $n \geq 1$, the stopping times τ_n are almost surely finite.

5.2. **Elementary properties of the branching process.** By construction of the process, 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 the time changes of a random walk with Bernoulli$(1-p)$ increments. Thus by the strong law of large numbers,

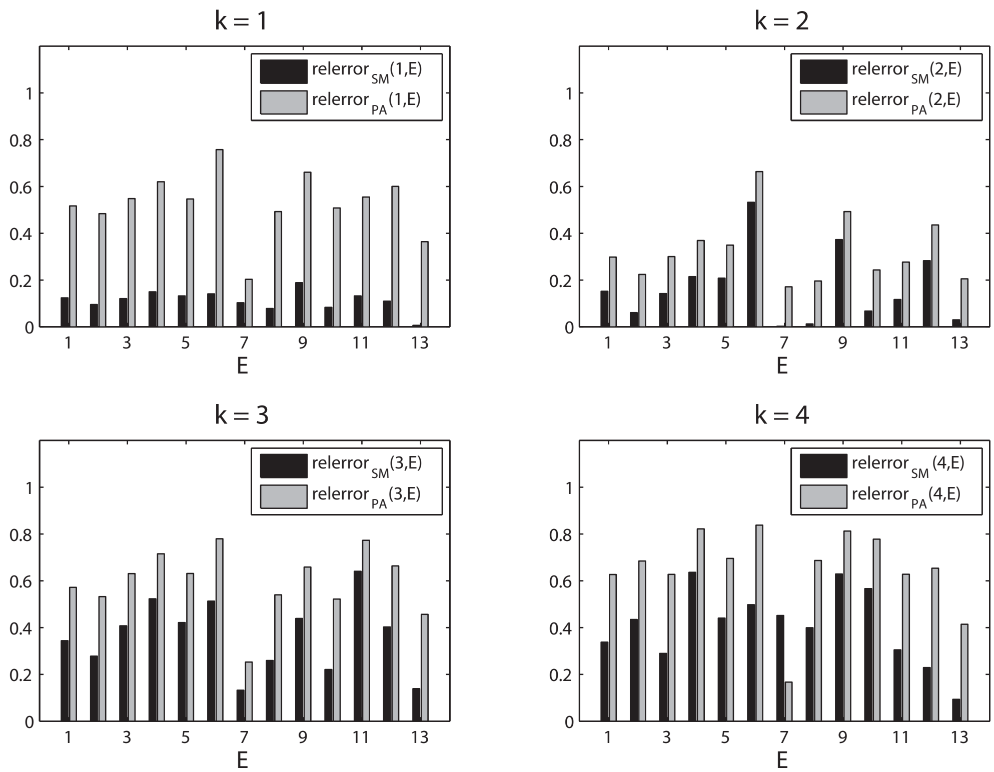

FIGURE 4.3. Plots of the relative errors of the degree distribution predictions of preferential attachment and the Superstar Model for 13 retweet graphs. The errors are plotted for degree $k = 1, 2, 3, 4$

we have

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

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

Definition 5.1 (Rate a Yule process)**.** Fix $a > 0$. A rate a Yule process is defined as a pure birth process $\mathsf{Yu}_a(\cdot)$ which starts with a single individual $\mathsf{Yu}_a(0) = 1$ with the rate of creating a new individual proportional to the number of present individuals in the population with

$$\mathbb{P}(\mathsf{Yu}_a(t + dt) - Yu_a(t) = 1 | \mathsf{Yu}_a(t)) = a\mathsf{Yu}_a(t)dt.$$

The Yule process is well studied and the next Lemma collects some of its standard properties (see [23], Section 2.5).

Lemma 5.2 (Yule process)**.**

(a) For any $t > 0$, $\mathsf{Yu}_a(t)$ has a geometric distribution with

$$\mathbb{P}(\mathsf{Yu}_a = k) = e^{-at}(1 - e^{-at})^{k-1}, \quad k \geq 1.$$

(b) The process $e^{-at}\mathsf{Yu}_a(t)$ is an $\mathbb{L}^2$ bounded martingale with respect to the natural filtration

of the process. Thus $e^{-at}\mathsf{Yu}_a(t) \xrightarrow{a.s.} W'$, *where* W' *has an exponential distribution with mean one.*

Now define the process

$$M(t) = e^{-(2-p)t}\left(|\mathsf{BP}(t)| + B(t)\right) \quad 0 \le t < \infty.$$

Proposition 5.3 (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 to a random variable* $M(t) \to W^*$ *almost surely and in* $\mathbb{L}^2$ *with* $\mathbb{E}(W^*) = 1$. *The random variable* $W^* > 0$ *with probability one. By* (5.2)

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

Proof Write $Z(t) = |\mathsf{BP}(t)|$ and $Y(t) = Z(t) + B(t)$ so that $M(t) = e^{-(2-p)t}Y(t)$. We shall denote $dM(t) = M(t+dt) - M(t)$. Then

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

Note that the processes $Z(t), B(t)$ are all counting process which increase by increments of one. For such processes, we shall use the infinitesimal notation $\mathbb{E}(dZ(t)|\mathsf{BP}(t)) = a(s)ds$ 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

$$\mathbb{P}\left(dZ(t) = 1 |\, \mathsf{BP}(t)\right) = \left(\sum_{v\in\mathcal{F}(t)} (c_B(v,t) + 1)\right) dt$$

where $c_B(v,t)$ denotes the number of blue children of vertex v at time t. The number of blue vertices can be written as $B(t) = \sum_{v\in\mathcal{F}(t)} c_v(b;t)$ since every blue vertex is an offspring of a unique vertex in $\mathsf{BP}(t)$ and is counted exactly once in this sum. Thus using the rate description, we get the conditional expectation

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

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 5.2 that for all $t \ge 0$, $\mathbb{E}(|\mathsf{BP}(t)|^2) < \infty$.

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 which happens with probability $1-p$. Thus we get

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

Combining we get

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

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

Let us next show that $M(\cdot)$ is an $\mathbb{L}^2$ bounded martingale. The process $Y^2(t+dt)$ can take values $(Y(t)+1)^2$ or $(Y(t)+2)^2$ at rate $pY(t)$ and $(1-p)Y(t)$ respectively. Thus we get

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

In particular the process $U(t)$ defined as

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

is a martingale. Taking expectations and noting that since $M(\cdot)$ is a martingale, this implies that $\mathbb{E}(M(s)) = 1$ for all s gives

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

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

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

Using equation (5.2) shows that $e^{-(2-p)s}|\mathsf{BP}(t)| \to W^*/(2-p) := W$. Now we only need to show W is strictly positive. First note that by $\mathbb{L}^2$ convergence, $\mathbb{E}(W^*) = 1$. This shows that $\mathbb{P}(W = 0) = r < 1$. 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 i^{th} child and its descendants. Then

$$e^{-(2-p)t}|\mathsf{BP}(t)| = \sum_{j=1}^{\infty} e^{-(2-p)\zeta_i}\left[e^{-(2-p)(t-\zeta_i)}|\mathsf{BP}_i(t-\zeta_i)|\right] \mathbb{1}\{\zeta_i \leq t\} + e^{-(2-p)t}.$$

Thus as $t \to \infty$ we have the distributional identity $W = \sum_{j=1}^{\infty} e^{-(2-p)\zeta_i} W_i$ where $\{W_i\}_{i\geq 1}$ are independent and identically distributed with the same distribution as W (independent of $\{\zeta_i\}_{i\geq 1}$). Thus

$$\mathbb{P}(W = 0) = \mathbb{P}(W_i = 0 \ \forall\ i \geq 1) = 0.$$

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. Also define the process

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

Lemma 5.4 (Offspring distribution properties)**.**

(a) Conditional on $\mathsf{BP}(\sigma_v)$, the process $c_B(v, \sigma_v + \cdot)$ has the same distribution as $\mathsf{Yu}_{1-p}(\cdot) - 1$. In particular $\mathbb{E}(c_B(v,t)) = e^{(1-p)t} - 1$.
(b) The process $M^*(t)$ is a martingale with respect to the filtration $\{\mathsf{BP}(\sigma_v + s) : s \geq 0\}$. In particular $\mathbb{E}(c_R(v, \sigma_v + t)) = \frac{p}{1-p}(e^{(1-p)t} - 1)$.

Proof. Part(a) is obvious from construction. To prove (b), note that $\mathbb{E}(dc_R(v, \sigma_v + t)|\mathsf{BP}(t + \sigma_v)) = p(c_B(v, \sigma_v + t) + 1)dt$. □

5.3. **Convergence for blue children proportions.** The equivalence between $\mathsf{BP}(\cdot)$ and the superstar model 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 $t \to \infty$.

Theorem 5.5. *Fix $k \geq 1$ and let $Z_{\geq k}(t)$ denote the number of vertices in $\mathsf{BP}(t)$ which 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 $t \to \infty$, where W^* is the martingale limit obtained in Proposition 5.3 and $p_{\geq k}(\infty)$ is defined by

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

Proof: The proof uses a variant of the "reproduction martingale" technique developed in [21]. The proof relies on two steps:

(a) Proving convergence of expectations of the desired quantities: Section 5.3.1.

(b) Bootstrapping this to a.s. convergence using law of large numbers: Section 5.3.2.

We setup some initial notation to carry out this program. Write $\xi = (\zeta_1, \zeta_2, \ldots,)$ for the offspring birth times of the root vertex v_1 (the offspring distribution of any vertex in $\mathsf{BP}(\cdot)$ is independent with the same distribution). For $t \geq 0$, let $\xi[0,t]$ denote the number of offspring in the interval $[0,t]$ and let $\mu[0,t] = \mathbb{E}(\xi[0,t])$ be the corresponding intensity measure. We start with a simple Lemma which will have profound consequences.

Lemma 5.6 (Renewal measure)**.** *Define $\alpha = 2 - p$. Then*

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

Thus the measure defined as $\mu_\alpha := e^{-\alpha t}\mu(dt)$ is a probability measure. Further this measure has expectation $\int_0^\infty t \mu_\alpha(dt) = 1$.

Proof: Recall that in Lemma 5.4 we used $c_B(v_1, t), c_R(v_1, t)$ to denote the number of red and blue children respectively of vertex v_1. Then $\mu([0,t]) = \mathbb{E}(c_R(v_1,t) + c_B(v_1,t))$. 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 $\mathbb{E}(c_B(v_1,t)), \mathbb{E}(c_R(v_1,t))$ from Lemma 5.4 completes the proof. The second assertion regarding the expectation follows similarly. $\square$

5.3.1. *Convergence of expectations.* The first step in the proof of Theorem 5.5 is convergence of expectations. This follows using standard renewal theory. However we will first need to setup notation that will allow us to use the linearity of expectations to derive a renewal equation.

Let us motivate an abstract definition of a *characteristic*. Fix some time $t > 0$. Suppose we are interested in the number of vertices with at least k blue children at this time. For any vertex $v \in \mathsf{BP}(\cdot)$, write σ_v for the time of birth of the vertex into $\mathsf{BP}(\cdot)$. Then conditional on $\mathsf{BP}(\sigma_v)$, the distribution of the number of blue children of vertex v by time $t - \sigma_v$ is $\mathsf{Yu}_{1-p}^{v}(t - \sigma_v) - 1$, where we construct a countable family of independent rate $1-p$ Yule processes $\mathsf{Yu}_{1-p}^{v}(\cdot)$ and use these to construct $\mathsf{BP}(\cdot)$ along with additional randomization for the red vertices. In particular, writing $Z_{\geq k}(t)$ for the number of vertices with degree at least k, this can be expressed as

$$Z_{\geq k}(t) = \sum_{v \in \mathsf{BP}(t)} \mathbb{1}\left\{[\mathsf{Yu}^v_{1-p}(t-\sigma_v) - 1] \geq k\right\}.$$

This motivates the following abstract construction. Let $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ be a bounded ($\sup_t \phi(t) < C$ for some non-random constant C) non-negative measurable stochastic process which depends only on the offspring distribution of a single vertex, often referred to as a characteristic, see e.g.[16]. Let $\phi^v(\cdot)$ be copies of this characteristic for each vertex $v \in \mathsf{BP}$. Finally 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:** $\phi(t) = 1$ gives $Z_\phi(t) = |\mathsf{BP}(t)|$.

(b) **Degree:** $\phi(t) = \mathbb{1}\{k \text{ or more blue children at time } t\}$ gives $Z_\phi(t) = Z_{\geq k}(t)$.

Fix any time $t > 0$. Conditioning on the offspring distribution of v_1, both of these characteristics satisfy the recursion

$$Z_\phi(t) = \phi^{v_1}(t) + \sum_{\zeta_i \leq t} Z^{(i)}_\phi(t - \zeta_i), \tag{5.5}$$

where $Z^{(i)}_\phi(\cdot) \stackrel{d}{=} Z_\phi(\cdot)$ and are independent. Taking expectations and writing $m_\phi(t) = \mathbb{E}(Z_\phi(t))$, these functions satisfy the renewal equation

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

Lemma 5.6 and renewal theory ([15]) now imply the next result.

Proposition 5.7. *For bounded characteristics, writing $\alpha = (2-p)$ we have*

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

Corollary 5.8. *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 $\phi(t) = \mathbb{1}\{k$ or more blue children at time $t\}$

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

Proof: The first assertion in the corollary is obvious. To prove the second assertion regarding the number of blue vertices, observe that the limit constant in Proposition 5.7 can be written as

$$\frac{1}{\alpha}\int_0^\infty \alpha e^{-\alpha s}\mathbb{E}(\mathbb{1}\{k \text{ or more blue children at time } s\})ds = \frac{1}{\alpha}\mathbb{P}(c_B(v_1, T) \geq k)$$

where T is an exponential random variable with mean α^{-1} that is independent of the blue offspring distribution $c_B(v_1, \cdot) = \mathsf{Yu}_{1-p}(\cdot) - 1$ where $\mathsf{Yu}_{1-p}(\cdot)$ is rate $1-p$ Yule process. The inter-arrival 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 $\mathbb{P}(c_B(v_1, T) > k) = \mathbb{P}(T > \sum_{j=0}^{k-1} X_j)$. One can check that the last probability equals $p_{\geq k}(\infty)$. $\square$

5.3.2. *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 [21] to analyze single type branching processes as well as in [18] to analyze times of first birth in generations. As before let $v_1, v_2, v_3, \ldots$ denote the order in which vertices appear and let $\tau_i = \sigma_{v_i}$ denote the times at which these vertices are born into the branching process $\mathsf{BP}(\cdot)$. Let $\xi^{(i)} = (\zeta_{v_i,1}, \zeta_{v_i,2}, \ldots)$ denote the offspring point process of v_i. Viewing $\xi^{(i)}$ as a random measure on $\mathbf{R}^+$, we get

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

For $m \geq 1$ let $\tilde{\mathcal{F}}_m$ be the sigma-algebra generated by vertices $\{v_1, \ldots, v_m\}$ and their offspring distribution point process (i.e. for $1 \leq i \leq m$, $\tilde{\mathcal{F}}_m$ has the type of v_i, times of birth as well as types of all the offspring). Define $\tilde{R}_0 = 1$ and

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

Let Γ_m be the set of the first m individuals born and **all** of their offspring. It is easy to check that

$$\tilde{R}_m = 1 + \sum_{v\in\Gamma_m} e^{-\alpha\sigma_v} - \sum_{j=1}^{m} e^{-\alpha\sigma_{v_j}}. \tag{5.6}$$

In particular $\tilde{R}_m > 0$ for all m. The next Lemma shows that the sequence $\left\{\tilde{R}_m\right\}_{m\geq 1}$ is much more.

Proposition 5.9 (Reproduction martingale)**.** *The sequence $\left\{\tilde{R}_m\right\}_{m\geq 1}$ is a non-negative martingale with respect to the filtration $\left\{\tilde{\mathcal{F}}_m\right\}_{m\geq 1}$. 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 5.6, $\mathbb{E}(\xi_\alpha^{(i)}) = \int_0^\infty e^{-\alpha t}\mu(dt) = 1$. Further $\sigma_{v_{m+1}}$ is $\tilde{\mathcal{F}}_m$ measurable while $\xi_\alpha^{(m+1)}$ is independent of $\tilde{\mathcal{F}}_m$. Thus one gets

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

Now assuming $\mathbb{E}([\xi_\alpha^{(i)}]^2) < \infty$, we see by the orthogonal increments of the martingale R_m that

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

Thus to check $\mathbb{L}^2$ boundedness it is enough to check that the right hand side is bounded. The following lemma accomplishes this.

Lemma 5.10.
(a) Assume $0 < p < 1$. Then $\mathbb{E}([\xi_\alpha]^2) < \infty$.
(b) For any m, $\mathbb{E}(\sum_{i=1}^m e^{-2\alpha\sigma_{v_i}}) \leq 1 + \alpha^{-1}$

Proof: To prove (a), we observe that $\xi_\alpha = \int_0^\infty \alpha e^{-\alpha t}\xi[0,t]dt = \mathbb{E}(\xi[0,T])$, where T is an exponential random variable with mean α^{-1} independent of ξ which is the offspring distribution of v_1. Thus it is enough to show $\mathbb{E}([\xi[0,T]]^2) < \infty$. Note that $\xi[0,T] = c_R(v_1,T) + c_B(v_1,T)$, i.e. 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 $T = t$ and noting by Lemma 5.2 that for fixed t, $\mathbb{E}(c_B^2(v_1,t)) \le Ce^{2(1-p)t}$ while for any t, conditional on $c_B(v_1,t)$, $c_R(v_1,t)$ is stochastically bounded by a Poisson random variable with rate $tc_B(v_1,t)$. Noting that $\alpha = 2-p$, we get

$$\mathbb{E}([\xi[0,T]]^2) \le C\int_0^\infty e^{-(2-p)t}\left(e^{2(1-p)t} + t^2e^{2(1-p)t}\right)dt < \infty.$$

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, since the rate of creation of new vertices is $|\mathsf{BP}(t)| + B(t)$ (see Proposition 5.3), one has

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

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. □

The next Theorem completes the proof of Theorem 5.5. Recall the limit constant $\tilde{m}_\phi(\infty)$ in Proposition 5.7.

Theorem 5.11 (Convergence of characteristics)**.** *For any bounded characteristic which satisfies the recursive decomposition in* (5.5) *one has*

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

Taking $\phi = 1$ and using Proposition 5.3 implies that $R_\infty = W$, the a.s. limit of the martingale $e^{-\alpha t}(|\mathsf{BP}(t)| + B(t))$.

Proof: A key role will be played by the martingale $\left\{\tilde{R}_n\right\}_{n\ge 0}$. Recall that this was a martingale with respect to the filtration $\{\mathcal{F}_m\}_{m\ge 0}$. We shall switch gears and now think about the process in continuous time. Define $I(t)$ as the set of individual born after time t whose mothers were born before time t and let

$$R_t = \sum_{x\in I(t)} e^{-\alpha\sigma_x} := \tilde{R}_{|\mathsf{BP}(t)|}, \qquad \{\mathcal{F}_t\}_{t\ge 0} := \left\{\tilde{\mathcal{F}}_{|\mathsf{BP}(t)|}\right\}_{t\ge 0}.$$

It is easy to check that R_t is an $\mathbb{L}^2$ bounded martingale with respect to this filtration and further $R_t \xrightarrow{a.s.} R_\infty$. For a fixed $c > 0$, define $I(t,c)$ as the set of vertices born after time $(t+c)$ whose mothers are born before time t and let $R_{t,c} = \sum_{x\in I(t,c)} e^{-\alpha\sigma_x}$. Obviously $R_{t,c} \le R_t$. Intuitively one should expect $R_{t,c}$ to be small for large c. The next Lemma quantifies this fact. First recall the random variable $\xi_\alpha = \int_0^\infty e^{-\alpha t}\xi(dt)$ and write $\xi_\alpha(c) = \int_c^\infty e^{-\alpha t}\xi(dt)$. Let $K(c) = \mathbb{E}(\xi_\alpha(c))$. It is easy to check that $K(c) \to 0$ as $c \to \infty$.

Proposition 5.12. *There exists a constant A such that for all $c > 0$*

$$\limsup_{t\to\infty} R_{t,c} \le K(c)W$$

where $W = \lim_{t\to\infty} e^{-\alpha t}|\mathsf{BP}(t)|$ from Proposition 5.3.

Proof: Without loss of generality we shall assume $t = k$ is an integer. The proof extends easily to general t. A key role is played by a strong law of large numbers, see [3] or [6] for a proof. This result was crucially used in [21] to prove convergence in the one type setting.

Lemma 5.13 (Strong law)**.** *Let $n_i, i = 1, 2, \ldots$ be a sequence of positive integers and let $X_{i,j}$ for $j = 1, 2, \ldots, n_i$ be a triangular array, independent for each fixed i and constructed on the same probability space. Suppose there exists a random variable $Y \geq 0$ with $\mathbb{E}(Y) < \infty$ such that $|X_{ij}|$ is stochastically dominated by Y. Further suppose that*

$$\liminf_{i\to\infty} \frac{n_{i+1}}{n_1 + \ldots + n_i} > 0. \tag{5.7}$$

Then

$$S_i = \frac{\sum_{j=1}^{n_i}(X_{ij} - \mathbb{E}(X_{ij}))}{n_i} \xrightarrow{a.s.} 0$$

as $i \to \infty$. Further assume the random variables are independent as i varies. The same is true of $\tilde{S}_k = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \mathbb{E}(X_{ij}))/(\sum_{i=1}^{k} n_i)$.

Proof of Proposition 5.12: Fix $t = k$, where k i an integer. By definition $R_{k,c}$ is made up contributions from all vertices u who are born after time $k + c$ whose mother $v = v(u)$ are in $\mathsf{BP}(k)$. Decomposing the sum $R_{k,c}$ according to the times of birth of this mother one has

$$R_{k,c} = \sum_{i=1}^{k} \sum_{v:\sigma_v\in[i-1,i)} e^{-\alpha\sigma_v} \int_{k+c-\sigma_v}^{\infty} e^{-\alpha s}\xi^v(ds).$$

Writing $\xi_\alpha^v(y) = \int_y^\infty e^{-\alpha t}\xi^v(dt)$ where $\xi^v(\cdot)$ is the offspring distribution point process of v, one immediately has

$$R_{k,c} \leq e^{-\alpha k} \sum_{i=1}^{k} \sum_{v:\sigma_v\in[i-1,i)} \xi_\alpha^v(c)$$

Each of these random variables are independent across different v and further are all stochastically bounded by the random variable $\xi_\alpha(c)$. Writing $n_i = \mathsf{BP}(i) - \mathsf{BP}(i-1)$, Prop 5.3 implies that the conditions in Lemma 5.13 are satisfied. Thus one has

$$e^{-\alpha k}\mathsf{BP}(k)\frac{\sum_{i=1}^{k}\sum_{v:\sigma_v\in[i-1,i)} \xi_\alpha^v(c)}{\mathsf{BP}(k)} \xrightarrow{a.s.} W\mathbb{E}(\xi_\alpha(c)).$$

This completes the proof. $\square$

Completing the proof of Theorem 5.11: Recall that we are dealing with bounded characteristics, i.e. $|\phi|_\infty < C$ for some constant C. Without loss of generality, let $C = 1$. We shall show that there exists a constant κ such that for all $\varepsilon > 0$,

$$\limsup_{t\to\infty} |e^{-\alpha t}Z_\phi(t) - \tilde{m}_\phi(\infty)| \leq \kappa\varepsilon(W + R_\infty). \tag{5.8}$$

Since this is true for any arbitrary ε, this completes the proof. Thus fix any $\varepsilon > 0$. First choose c large such that the function arising in the bound of Proposition 5.12 $K(c) < \varepsilon$. Next, define ϕ_s as the truncated characteristic

$$\phi_s(u) = \begin{cases} \phi(u), & u \leq s \\ 0, & u > s \end{cases} \tag{5.9}$$

This characteristic is zero for any vertices who have been alive for more that s, so we can view it as a characteristic for "young" vertices. The limit constant for this characteristic by Proposition 5.7 is

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

Here ϕ is the original characteristic. If we write $\phi' = \phi - \phi_s$, we can view ϕ' as the characteristic for "old" vertices. With this notation we have $Z_\phi(u) = Z_{\phi_s}(u) + Z_{\phi'}(u)$.

Define $\tilde{m}_{\phi_s}(u) = e^{-\alpha u}\mathbb{E}(Z_{\phi_s}(u))$. Now choose choose s large enough such that $s > c$ and 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 $Z_{\phi_s}(\cdot)$, which is the branching process counted according to the truncated characteristic. We first observe that since $\phi_s(u) = 0$ when $u > s$, for any $t > s$, vertices born before time $t - s$ (old vertices) do not contribute to $Z_{\phi_s}(t)$. Thus we can write

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

where $Z^x_{\phi_s}(t-\sigma_x)$ are the contributions to $Z_{\phi_s}(t)$ by the descendants of a vertex x born in the interval $[t-s,t]$ whose mother belongs to $\mathsf{BP}(t-s)$. Let $\mathcal{N}(t,c) = I(t) \setminus I(t,c)$, i.e. the set of individuals born in the interval $[t, t+c]$ to mothers who were born before time t. Then we can decompose the difference as a telescoping sum:

$$e^{-\alpha t}Z_\phi(t) - m_\phi(\infty) := E_1(t) + E_2(t) + E_3(t) + E_4(t) + E_5(t). \tag{5.10}$$

Here:

(a) $E_1(t)$ is defined as

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

Observe that for $E_1(t)$, the only vertices which 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 Prop 5.3, one has $\limsup_{t\to\infty} E_1(t) \le e^{-\alpha s}W \le \varepsilon W$ by choice of s.

(b) $E_2(t)$ is defined as

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

For $E_2(t)$, $\mathcal{N}(t-s,c)$ consists of all children of mothers in $\mathsf{BP}(t-s)$ born in the interval $[t-s, t-s+c]$. Since 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 $\liminf_{t\to\infty} |\mathcal{N}(t-s,c)|/|\mathsf{BP}(t-s)| \ge c$. Further 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 $Z_{\phi_s}(c)$. Similar to the proof of Prop 5.12, using Lemma 5.13 one can show that $\limsup_{t\to\infty} |E_2(t)| \to 0$ a.s. We omit the details.

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

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

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

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

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

For $E_4(t)$, we have $|\left(\sum_{x\in\mathcal{N}(t-s,c)} e^{-\alpha\sigma_x} - R_{t-s}\right)| = R_{t-s,c}$. Thus $\limsup_{t\to\infty} E_4(t) \leq \tilde{m}_{\phi_s}(\infty)K(c)W \leq \tilde{m}_{\phi_s}(\infty)\varepsilon W$ by choice of c and using Proposition 5.12 for the asymptotics of $R_{t,c}$.

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

Combining all these bounds, one finally arrives at

$$\limsup_{t\to\infty} |e^{-\alpha t}Z_\phi(t) - \tilde{m}_\phi(\infty)| \leq \varepsilon(W + \tilde{m}_{\phi_s}(\infty)R_\infty).$$

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

5.4. **Time of first birth asymptotics.** For a rooted tree with root ρ, there is a natural notion of a generation of a vertex v, which is 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 $\mathsf{BP}(t)$. For each fixed k, we shall then define a sequence of stopping times $\mathrm{Bir}(k)$ representing the first time an individual in modified generation k is born into the process $\mathsf{BP}(\cdot)$. We shall study asymptotics of $\mathrm{Bir}(k)$ as $k\to\infty$. In the next Section we shall show how these asymptotics result in 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 between v and 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

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

This is equivalent to 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. The next theorem proves asymptotics for these times.

Theorem 5.14. *Let $W(\cdot)$ be the Lambert function. We have*

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

Proof of Theorem 5.14: 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 a $\mathsf{Yu}_{1-p}(\cdot)-1$ offspring distribution. Further let

$$\mathrm{Bir}^*(k) = \inf\left\{t : \exists\, v\in BP_b^{v_1}(t),\ G(v)=k\right\}.$$

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 $\mathrm{Bir}^*(k)/k \xrightarrow{a.s.} \beta$. Here β can be derived as follows. Write μ_b for the expected intensity measure of the blue offspring distribution, i.e. $\mu_b([0,t]) = \mathbb{E}(c_B[v_1,t]) = e^{(1-p)t} - 1$ from Lemma 5.4. For $\theta > 0$, let $\phi(\theta) := \mathbb{E}(\int_0^\infty e^{-\theta t} c_B(v_1, dt))$. 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

$$\Lambda(a) := \inf \left\{ \phi(\theta) e^{\theta a} : \theta \geq 1-p \right\} = (1-p) a e^{(1-p)a+1}. \tag{5.11}$$

Then the limit constant β is derived as

$$\beta = \sup \left\{ a > 0 : \Lambda(a) < 1 \right\}. \tag{5.12}$$

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

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

This gives an upper bound in Theorem 5.14. Lemma 5.15 proves a lower bound and completes the proof.

Lemma 5.15. *Fix any $\varepsilon > 0$ and let $\beta = W(1/e)/(1-p)$ be the limit constant. Then*

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

Thus one has $\liminf_{l\to\infty} Bir(l)/l \geq \beta$ 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 distribution $\mathsf{Yu}_{1-p}(\cdot) - 1$.

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 $\mathrm{Bir}^*(v,l) - \sigma_v \overset{d}{=} \mathrm{Bir}^*(l)$, conditional on $\mathsf{BP}(\sigma_v)$ one has

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

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

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

One of the main bounds of Kingman ([18], Theorem 1) is $\mathbb{E}[e^{-\theta}\mathrm{Bir}^*(l)] \le (\phi(\theta))^l$. Thus we get

$$\mathbb{P}(\mathrm{Bir}^*(l) < (1-\varepsilon)\beta l - s) \le [\phi(\theta)e^{\theta(1-\varepsilon)\beta}]^l e^{-\theta s}. \tag{5.13}$$

By the definition of β,

$$\Lambda_\varepsilon := \Lambda(\beta(1-\varepsilon)) := \inf\left\{\phi(\theta)e^{\theta(1-\varepsilon)\beta} : \theta > 1-p\right\} < 1.$$

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

$$\mathbb{P}(\mathrm{Bir}^*(l) < (1-\varepsilon)\beta l - s) \le [\Lambda_\varepsilon]^l e^{-\theta_\varepsilon s}. \tag{5.14}$$

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 $t_l(\varepsilon)$. 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

$$\begin{aligned}\mathbb{E}(N_l^\varepsilon) &\le \int_0^{t_l(\varepsilon)} [\Lambda_\varepsilon]^l d\mathbb{E}(R(s)) \quad \text{using Eqn. (5.14)},\\ &= p[\Lambda_\varepsilon]^l \int_0^{t_\varepsilon(l)} e^{-(\theta_\varepsilon - q)s} ds \text{ using Lemma 5.4.}\end{aligned}$$

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

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

$\square$

6. Equivalence between the branching process and the superstar model

We start with an informal description of the connection between the Superstar Model and the branching process $\mathsf{BP}(\cdot)$. We connect vertex v_1, which is the initial progenitor of $\mathsf{BP}(\cdot)$, to the superstar v_0 (which does not play a role in the evolution of $\mathsf{BP}(\cdot)$) in G_2. All the red vertices in the process $\mathsf{BP}(\cdot)$ correspond to the neighbors of the superstar v_0. The true degree of a (non-superstar) 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 it's ancestor. By elementary properties of the exponential distribution, the dynamics of $\mathsf{BP}(\cdot)$ imply that each new vertex which is born is red (connected to the superstar v_0) with probability p, else with probability $q = 1-p$ is blue and connected to any other vertex with probability proportional to it's current degree, increasing the degree of this chosen vertex by one. This is nothing but the Superstar Model.

Formally our surgery will take the random tree $\mathsf{BP}(\tau_n)$ and modify it to get an $n+1$-vertex tree $\mathcal{S}_n$ which 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.

As before we label the vertices of $\mathsf{BP}(\tau_n)$ by $\{v_1, v_2, \ldots, v_n\}$ in order of their birth and then we add a new vertex v_0 to this set to give us the vertex set for G_{n+1}. One can anticipate that v_0 will be our superstar.

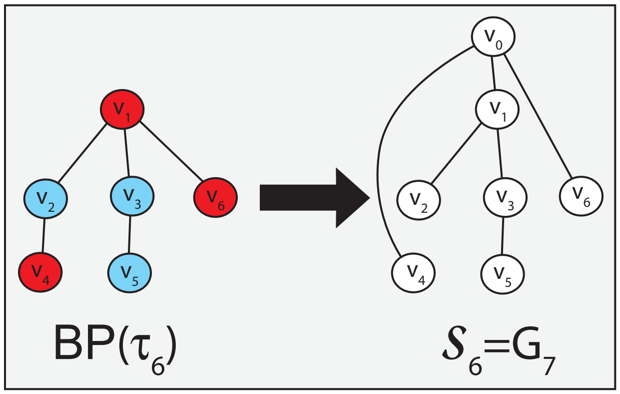

FIGURE 6.1. The surgery passing from $\mathsf{BP}(\tau_n)$ to $\mathcal{S}_{n+1}$ and G_{n+1} for $n = 6$.

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 6.1.

Proposition 6.1 (Equivalence from surgery operation)**.** *The tree $\mathcal{S}_n$ viewed as a tree with vertices without colors has the same distribution as the Superstar Model G_{n+1}. In fact the process $\{\mathcal{S}_n\}_{n\geq 1}$ has the same distribution as $\{G_{n+1}\}_{n\geq 1}$.*

Proof: We shall prove this by induction. 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 ancestor. The ancestor of all the red individuals is the superstar v_0 while the ancestors 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 non-superstar vertex in $\mathcal{S}_n$ is the number of blue children it possesses plus one for the edge connecting the vertex to it's ancestor 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

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

Thus a new vertex 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 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 proof we shall assume G_{n+1} is constructed through this surgery process and suppress $\mathcal{S}_n$.

7. Proofs of the main results

Let us now prove all the main results by using the equivalence created by the surgery operation and the proven results on $\mathsf{BP}(\cdot)$ in Section 5. We record the following fact about the asymptotics for the stopping times τ_n.

Lemma 7.1 (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 5.3 proves that $|\mathsf{BP}(t)|e^{-(2-p)t} \xrightarrow{a.s.} W$. Thus $ne^{-(2-p)\tau_n} \xrightarrow{a.s.} W$. □

7.1. **Proof of the Superstar strong law.** By the surgery operation, the degree of the superstar is given by $R(\tau_n)$, the total number of red vertices. Equation (5.2) shows that the number of blue vertices satisfies $B(\tau_n)/|\mathsf{BP}(\tau_n)| \xrightarrow{a.s.} 1-p$. Thus $R(\tau_n)/|\mathsf{BP}(\tau_n)| \xrightarrow{a.s.} p$. This completes the proof. □

7.2. **Proof of the degree distribution strong law.** Since G_{n+1} is a connected tree, every vertex has degree at least one. Recall that $c_B(v,t)$ denoted 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 non-superstar vertex

$$\deg(v, G_{n+1}) = c_B(v,\tau_n) + 1. \tag{7.1}$$

Fixing $k \geq 0$, the number of non-superstar vertices with degree exactly $k+1$ is the same as the number of number of vertices in $\mathsf{BP}(\tau_n)$ which have exactly k blue children. Recall that we used $Z_{\geq k}(t)$ for the number of vertices in $\mathsf{BP}(t)$ which have at least k blue children. Proposition 5.3, showed that the total number of vertices $|\mathsf{BP}(t)|$ satisfies $e^{-(2-p)t}|\mathsf{BP}(t)| \xrightarrow{a.s.} W^*/(2-p)$. Theorem 5.5 showed that

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

Thus writing $p_{\geq k}(t) = Z_{\geq k}(t)/\mathsf{BP}(t)$ for the proportion of vertices with degree k, Theorem 5.5 implies one has

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

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

7.3. **Proof of maximal degree asymptotics.** The aim of this is to prove Theorem 2.4. We wish to analyze the maximal non-superstar degree which we wrote as

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

The plan will be as follows: we will first prove the simpler assertion of convergence of the degree of vertex v_k for fixed $k \geq 1$. 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 $k \geq 1$. Recall from (7.1) 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 5.2,

$$\frac{c_B(v_k, t)}{e^{(1-p)(t-\tau_k)}} \xrightarrow{a.s.} W_k', \tag{7.2}$$

where W_k' is an exponential random variable with mean one. By Proposition 5.3, $|\mathsf{BP}(t)|/e^{(2-p)t} \xrightarrow{a.s.} W$. Write $\gamma = (1-p)/(2-p)$ and let $\Delta_k = e^{-(1-p)\tau_k} W' W^{-\gamma}$. Then we have

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

□

Now let us prove the convergence of the maximal non-superstar degree Υ_n. Fix $L > 0$ and let

$$\tilde{M}_n[0, L] := \max \left\{ \deg(v_k, G_{n+1}) : \tau_k \leq L \right\}. \tag{7.3}$$

In 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 $k \geq 1$ implies the next result.

Lemma 7.2 (Convergence near the root)**.** *Fix any $L > 0$. Then there exists a random variable $\Delta^*[0, L] > 0$ such that*

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

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. First we shall need to setup some notation. Recall that by asymptotics for the stopping times τ_n in Lemma 7.1, given any $\varepsilon > 0$, we can choose $K_\varepsilon > 0$ such that

$$\limsup_{n \to \infty} \mathbb{P}\left(\left| \tau_n - \frac{1}{2-p} \log n \right| > K_\varepsilon \right) \leq \varepsilon. \tag{7.4}$$

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

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

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

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

Using Lemma 7.2 now shows that there exists a random variable Δ^* such that $\Upsilon_n/n^\gamma \xrightarrow{a.s.} \Delta^*$, and this completes the proof of Theorem 2.4.

Proof of Lemma 7.3: For ease of notation, write $t_n^- = (2-p)^{-1}\log n - K$ and $t_n^+ = (2-p)^{-1}\log n + K$. 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

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

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

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

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

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

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

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

Without loss of generality, we shall assume L_ε, t_n^+ are all 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

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

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

$$\mathbb{P}\left(M_{[L,t_n^+]}(t_n^+) > Cn^\gamma\right) \leq \sum_{m=L}^{t_n^+-1} \mathbb{E}(|\mathsf{BP}[m,m+1]|)\mathbb{P}(\mathrm{Yu}_q(t_n^+ - m) > Cn^\gamma).$$

Now $\mathbb{E}(\mathsf{BP}[m,m+1]) \leq \mathbb{E}(|\mathsf{BP}(m+1)|)$. By Proposition 5.3, $\mathbb{E}(|\mathsf{BP}(t)|) \leq e^{(2-p)t}$. Further by Lemma 5.2, for fixed time s, a rate $1-p$ Yule process has a geometric distribution with parameter $e^{-(1-p)s}$. Thus we have

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

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.

□

7.4. **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 5.4 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 $r(v)$ which 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

$$\mathcal{H}(G_{n+1}) = \max\left\{d(v) + 1 : v \in \mathsf{BP}(\tau_n)\right\}. \tag{7.5}$$

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

$$\mathrm{Bir}(\mathcal{H}(G_{n+1}) - 1) \le \tau_n \le \mathrm{Bir}(\mathcal{H}(G_{n+1})). \tag{7.6}$$

Now recall that Theorem 5.14 showed that the stopping times $\mathrm{Bir}(k)$ satisfy $\mathrm{Bir}(k)/k \xrightarrow{a.s.} W(1/e)/1 - p$ as $k \to \infty$. Dividing (7.6) throughout by $\mathcal{H}(G_{n+1})$ we have

$$\frac{\mathrm{Bir}(\mathcal{H}(G_{n+1}) - 1)}{\mathcal{H}(G_{n+1})} \xrightarrow{a.s.} \frac{W(1/e)}{1-p}, \qquad \frac{\tau_n}{\log n} \xrightarrow{a.s.} \frac{1}{2-p}.$$

Here the first assertion follows by Theorem 5.14 while the second assertion follows from Lemma 7.1 which described asymptotics for the stopping times τ_n. Rearranging shows that

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

This completes the proof. □

Acknowledgements

SB has been partially supported by NSF-DMS grant 1105581 and would like to thank the hospitality of the Statistics department at Wharton where this work commenced.

References

[1] Data courtesy Microsoft Research Cambridge, August 2010.

[2] Twitter developers. https://dev.twitter.com/docs/api/1.1/get/statuses/firehose, August 2012.

[3] S. Asmussen and T.G. Kurtz. Necessary and sufficient conditions for complete convergence in the law of large numbers. *The Annals of Probability*, pages 176–182, 1980.

[4] S. Asur, B.A. Huberman, G. Szabo, and C. Wang. Trends in social media: Persistence and decay. *AAAI Conference on Weblogs and Social Media*, 2011.

[5] K.B. Athreya, A.P. Ghosh, and S. Sethuraman. Growth of preferential attachment random graphs via continuous-time branching processes. In *Proceedings of the Indian Academy of Sciences-Mathematical Sciences*, volume 118, pages 473–494. Indian Academy of Sciences, 2008.
[6] K.B. Athreya and N. Kaplan. Convergence of the age distribution in the one-dimensional supercritical age-dependent branching process. *The Annals of Probability*, 4(1):38–50, 1976.
[7] A.L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
[8] B. Bollobás and O.M. Riordan. Mathematical results on scale-free random graphs. *Handbook of graphs and networks*, pages 1–34, 2003.
[9] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Struct. Algorithms*, 18(3):279–290, May 2001.
[10] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring user influence in Twitter: the million follower fallacy. *AAAI Conference on Weblogs and Social Media*, 2010.
[11] C. Cooper and A. Frieze. A general model of web graphs. *Random Structures & Algorithms*, 22(3):311–335, 2003.
[12] M. Deijfen. Random networks with preferential growth and vertex death. *Journal of Applied Probability*, 47(4):1150–1163, 2010.
[13] S.N. Dorogovtsev and J.F.F. Mendes. Evolution of networks. *Advances in physics*, 51(4):1079–1187, 2002.
[14] Rick Durrett. *Random graph dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2007.
[15] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.
[16] Peter Jagers. *Branching processes with biological applications*. Wiley-Interscience [John Wiley & Sons], London, 1975. Wiley Series in Probability and Mathematical Statistics—Applied Probability and Statistics.
[17] Peter Jagers and Olle Nerman. The asymptotic composition of supercritical multi-type branching populations. In *Séminaire de Probabilités, XXX*, volume 1626 of *Lecture Notes in Math.*, pages 40–54. Springer, Berlin, 1996.
[18] J. F. C. Kingman. The first birth problem for an age-dependent branching process. *Ann. Probability*, 3(5):790–801, 1975.
[19] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? *Proc. WWW*, 2010.
[20] T.F. Móri. Degree distribution nearby the origin of a preferential attachment graph. *Electron. Comm. Probab*, 12:276–282, 2007.
[21] Olle Nerman. On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrsch. Verw. Gebiete*, 57(3):365–395, 1981.
[22] M.E.J. Newman. The structure and function of complex networks. *SIAM review*, pages 167–256, 2003.
[23] J. R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original.
[24] Boris Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures & Algorithms*, 5(2):337–347, 1994.
[25] A. Rudas, B. Tóth, and B. Valkó. Random trees and general branching processes. *Random Structures & Algorithms*, 31(2):186–202, 2007.
[26] Anna Rudas and Bálint Tóth. Random tree growth with branching processes—a survey. In *Handbook of large-scale random networks*, volume 18 of *Bolyai Soc. Math. Stud.*, pages 171–202. Springer, Berlin, 2009.
[27] C. Wiuf, M. Brameier, O. Hagberg, and M.P.H. Stumpf. A likelihood approach to analysis of network data. *Proceedings of the National Academy of Sciences*, 103(20):7566–7570, 2006.

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”.