Discrete small world networks

A. D. Barbour*
Applied Mathematics
University of Zürich
CH - 8057 Zürich
and
G. Reinert†
Department of Statistics
University of Oxford
UK - Oxford OX1 3TG

Abstract

Small world models are networks consisting of many local links and fewer long range 'shortcuts', used to model networks with a high degree of local clustering but relatively small diameter. Here, we concern ourselves with the distribution of typical inter-point network distances. We establish approximations to the distribution of the graph distance in a discrete ring network with extra random links, and compare the results to those for simpler models, in which the extra links have zero length and the ring is continuous.

*ADB was supported in part by Schweizerischer Nationalfonds Projekt Nr. 20-67909.02.
†GR was supported in part by EPSRC grant no. GR/R52183/01.

1 Introduction

There are many variants of the mathematical model introduced by Watts and Strogatz [15] to describe the "small–world" networks popular in the social sciences; one of them, the great circle model of Ball *et. al.* [4], actually precedes [15]. See [1] for a recent overview, as well as the books [5] and [8]. A typical description is as follows. Starting from a ring lattice with L vertices, each vertex is connected to all of its neighbours within distance k by an undirected edge. Then a number of shortcuts are added between randomly chosen pairs of sites. Interest centres on the statistics of the shortest distance between two (randomly chosen) vertices, when shortcuts are taken to have length zero.

Newman, Moore and Watts [12], [13] proposed an idealized version, in which the lattice is replaced by a circle and distance along the circle is the usual arc length, shortcuts now being added between random pairs of uniformly distributed points. Within their [NMW] model, they made a heuristic computation of the mean distance between a randomly chosen pair of points. Then Barbour and Reinert [7] proved an asymptotic approximation for the distribution of this distance as the mean number $L\rho$ of shortcuts tends to infinity; the parameter ρ describes the average intensity of end points of shortcuts around the circle. In this paper, we move from the continuous model back to a genuinely discrete model, in which the ring lattice consists of exactly L vertices, each with connections to the k nearest neighbours on either side, but in which the random shortcuts, being edges of the graph, are taken to have length 1; thus distance becomes the usual graph distance between vertices. However, this model is rather complicated to analyze, so we first present a simpler version, in which time runs in discrete steps, but the process still lives on the continuous circle, and which serves to illustrate the main qualitative differences between discrete and continuous models. This intermediate model would be reasonable for describing the spread of a simple epidemic, when the incubation time of the disease is a fixed value, and the infectious period is very short in comparison. In each of these more complicated models, we also show that the approximation derived for the [NMW] model gives a reasonable approximation to the distribution of inter-point distances, provided that ρ (or its equivalent) is small; here, the error in Kolmogorov distance is of order $O(\rho^{\frac{1}{3}}\log(\frac{1}{\rho}))$, although the distribution functions are only $O(\rho)$ apart in the bulk of the distribution.

2 The continuous circle model for discrete time

In this section, we consider the continuous model of [7], which consists of a circle C of circumference L, to which are added a Poisson $\mathrm{Po}\,(L\rho/2)$ number of uniform and independent random chords, but now with a new measure of distance between points P and Q. This distance is the minimum of $d(\gamma)$ over paths γ along the graph between P and Q, where, if γ consists of s arcs of lengths $l_1, \ldots, l_s$ connected by shortcuts, then $d(\gamma) := \sum_{r=1}^{s} \lceil l_r \rceil$, where, as usual, $\lceil l \rceil$ denotes the smallest integer $m \geq l$; shortcuts make no contribution to the distance. We are interested in asymptotics as $L\rho \to \infty$, and so assume throughout that $L\rho > 1$.

We begin with a dynamic realization of the network, which describes, for each $n \geq 0$, the set of points $R(n) \subset C$ that can be reached from a given point P within time n, where time corresponds to the $d(\cdot)$ distance along paths. Pick Poisson $\mathrm{Po}\,(L\rho)$ uniformly and independently distributed 'potential' chords of the circle C; such a chord is an unordered pair of independent and uniformly distributed random points of C. Label one point of each pair with 1 and the other with 2, making the choices equiprobably, independently of everything else. We call the set of label 1 points Q, and, for each $q \in Q$, we let $q' = q'(q)$ denote the label 2 end point. Our construction realizes a random subset of these potential chords as shortcuts. We start by taking $R(0) = \{P\}$ and $B(0) = 1$, and let time increase in integer steps. $R(n)$ then consists of a union of $B(n)$ intervals of C, each of which is increased by unit length at each end point at time $n+1$, but with the rule that overlapping intervals are merged into a single interval; this defines a new union of $B'(n+1)$ intervals $R'(n+1)$; note that $B'(n+1)$ may be less than $B(n)$.

Now define $\partial R(n+1) := R'(n+1) \setminus R(n)$. Whenever $\partial R(n+1) \cap Q$ is not empty — that is, whenever $\partial R(n+1)$ includes label 1 points — then, for each $q \in \partial R(n+1) \cap Q$, we accept the chord $\{q, q'\}$ if $q' = q'(q) \notin R'(n+1)$ (that is, if the chord would reach beyond the cluster $R'(n+1)$), we reject it if $q' \in R(n)$, and we accept the chord $\{q, q'\}$ with probability $1/2$ if $q' \in \partial R(n+1)$, independently of all else. Letting $Q(n+1) := \{q' : \{q, q'\} \text{ newly accepted}\}$, take $R(n+1) = R'(n+1) \cup Q(n+1)$ and set $B(n+1) = B'(n+1) + |Q(n+1)|$. Note that $B(n+1)$ may be either larger or smaller than $B(n)$, and that $B_{\lceil L/2 \rceil} = 1$ a.s.

After at most $\lceil L/2 \rceil$ time steps, each of the potential chords has been either accepted or rejected independently with probability $1/2$, because of auxiliary randomization for those chords such that $\{q, q'\} \in \partial R(n)$ for some n, and because of the random labelling of the end points of the chords for the remainder. Hence this construction does indeed lead to $\mathrm{Po}\,(L\rho/2)$ independent uniform chords of C.

For our analysis, as in [7], we define a second process $S(n)$, starting from the same P and the same set of potential chords, and with the same unit growth per time step. The differences are that *every* potential chord is included, so that no thinning takes place, and, additionally, whenever two intervals intersect, they continue to grow, overlapping one another, and each continues to generate further chords according to a Poisson process of rate ρ. This pure growth process $S(n)$ agrees with the original construction during the initial development with high probability, until S has grown enough that overlap becomes likely; its advantage is that it has a branching structure, and is thus much more easily analysed. We denote its length at time n by $s(n) \geq r(n)$, overlaps now being counted according to multiplicity, and the number of intervals by $M(n) \geq B(n)$. Then $M(n)$ is just a pure birth chain with offspring distribution $1 + \mathrm{Po}\,(2\rho)$, so that $\mathbf{E}M(n) = (1+2\rho)^n$, and the total length of the $M(n)$ intervals is given by

$$s(n) = 2\sum_{k=0}^{n-1} M(k),$$

so that

$$\mathbf{E}s(n) = \rho^{-1}((1+2\rho)^n - 1).$$

Furthermore,

$$W(n) := (1+2\rho)^{-n} M(n)$$

forms a square integrable martingale, so that $(1+2\rho)^{-n} M(n) \to W_\rho$ a.s. for some W_ρ such that $W_\rho > 0$ a.s. and $\mathbf{E} W_\rho = 1$. Hence also $(1+2\rho)^{-n} s(n) \to \rho^{-1} W_\rho$ a.s. and $\frac{s(n)}{M(n)} \to \rho^{-1}$ a.s.. Note also that $\operatorname{Var} W(n) \le 1$.

Our strategy is to pick a starting point P, and run both constructions up to an integer time τ_r, chosen in such a way that $R(n)$ and $S(n)$ are (almost) the same for $n \le \tau_r$. Pick

$$n_0 = \left\lfloor \frac{\log(L\rho)}{2\log(1+2\rho)} \right\rfloor,$$

where $\lfloor x \rfloor$ denotes the largest integer no greater than x, and let

$$\phi_0 := \phi_0(L,\rho) = (L\rho)^{-1/2}(1+2\rho)^{n_0},$$

so that $(1+2\rho)^{n_0} = \phi_0\sqrt{L\rho}$ and $(1+2\rho)^{-1} \le \phi_0 \le 1$; note that $\phi_0 \approx 1$ if ρ is small. Now let $\tau_r = n_0 + r$, and assume that

$$|r| \le \frac{1}{6\log(1+2\rho)} \log(L\rho), \tag{2.1}$$

implying in particular that $\tau_r \le \frac{2\log(L\rho)}{3\log(1+2\rho)}$. Then, writing $R_r = R(\tau_r), S_r = S(\tau_r), M_r = M(\tau_r)$, and $s_r = s(\tau_r)$, we have

$$\mathbf{E} M_r = \phi_0\sqrt{L\rho}(1+2\rho)^r$$

and

$$\mathbf{E} s_r = \rho^{-1}(\phi_0\sqrt{L\rho}(1+2\rho)^r - 1).$$

Next, independently and uniformly, we pick a second point $P' \in C$, and a second set of potential chords, Q', and run both constructions for time $\tau_{r'}$, where r' also satisfies (2.1), yielding $R'_{r'}, S'_{r'}, M'_{r'} =: N_{r'}$ and $s'_{r'} =: u_{r'}$. Then, at least for small ρ, there are about $\phi_0^2 L\rho(1+2\rho)^{r+r'}$ pairs of intervals, with one in S_r and the other in $S'_{r'}$, and each is of typical length ρ^{-1}, so that the expected number of intersecting pairs of intervals is about

$$\frac{2}{L\rho}\phi_0^2 L\rho(1+2\rho)^{r+r'} = 2\phi_0^2(1+2\rho)^{r+r'},$$

which, in the chosen range of r, r', grows from almost nothing to the typically large value $2\phi_0^2(L\rho)^{1/3}$. For later use, label the intervals in S_r as $I_1, \ldots, I_{M_r}$, and the intervals in $S'_{r'}$ as $J_1, \ldots, J_{N_{r'}}$; then we can write the number $\widehat{V}_{r,r'}$ of intersecting pairs of intervals as

$$\widehat{V}_{r,r'} = \sum_{i=1}^{M_r} \sum_{j=1}^{N_{r'}} X_{ij}, \tag{2.2}$$

where

$$X_{ij} = \mathbf{1}\{I_i \cap J_j \neq \emptyset\}. \tag{2.3}$$

Now the probability that $\widehat{V}_{r,r'} = 0$ is the same as when the construction for S' uses the original set Q of potential chords, because of the independence of Poisson processes on disjoint subsets; the event $\widehat{V}_{r,r'} = 0$ indicates that the two processes have no intersecting pairs of intervals when stopped at the times τ_r, $\tau_{r'}$, and thus use disjoint sets of chords. Furthermore, we can show that the event $\widehat{V}_{r,r'} = 0$ is with high probability the same as the event $V_{r,r'} = 0$, where $V_{r,r'}$ is the number of intersections of $R(r)$ and $R'(r')$. Finally, if $R(r)$ and $R'(r')$ have no intersections, then the "small worlds" distance between P and P' is more than

$$\tau_r + \tau_{r'} = 2n_0 + r + r'.$$

Hence we have solved the problem if we can find a good approximation to the probability that $\widehat{V}_{r,r'} = 0$; this we do by showing that $\widehat{V}_{r,r'}$ approximately has a mixed Poisson distribution, and by identifying the mixture distribution. We usually take $r = r'$ or $r = r' + 1$, the latter to allow for the possibility of the number of steps in the shortest path being odd.

After this preparation, we are in a position to summarize our main results. These are treated in more detail in the next section, in Theorem 3.9, Corollary 3.10 and Theorem 3.15. We let D denote the small worlds distance between a randomly chosen pair of points P and Q on C, so that, as above,

$$\mathbf{P}[D > 2n_0 + r + r'] = \mathbf{P}[V_{r,r'} = 0].$$

The following theorem approximates the distribution of D by that of another random variable D^*, whose distribution is more accessible; in this theorem, ρ and the derived quantities ϕ_0, n_0, N_0 and x_0 all implicitly depend on L, as does the distribution of D^*.

Theorem 2.1 *Let Δ denote a random variable on the integers with distribution given by*

$$\mathbf{P}[\Delta > x] = \mathbf{E}\{e^{-2\phi_0^2(1+2\rho)^x W_\rho W'_\rho}\}, \quad x \in \mathbf{Z},$$

and set $D^* = \Delta + 2n_0$. If $L\rho \to \infty$ and $\rho = \rho(L) = O(L^\beta)$, with $\beta < 4/31$, then

$$d_{TV}(\mathcal{L}(D), \mathcal{L}(D^*)) \to 0 \quad as \quad L \to \infty.$$

1. *If ρ is large, let N_0 be such that $(1+2\rho)^{N_0} \le L\rho < (1+2\rho)^{N_0+1}$, and define $\alpha \in [0,1)$ to be such that $L\rho = (1+2\rho)^{N_0+\alpha}$; then, with $x_0 = N_0 - 2n_0 + 1$,*

$$\begin{aligned} \mathbf{P}[\Delta \ge x_0] &\ge 1 - 2(1+2\rho)^{-\alpha}; \\ \mathbf{P}[\Delta \ge x_0 + 1] &= O\left((1+2\rho)^{-1+\alpha}\log(1+\rho)\right), \end{aligned}$$

   so that Δ concentrates almost all its mass on x_0, unless α is very close to 1.

2. *If $\rho \to 0$, the distribution of $\rho\Delta$ approaches that of the random variable T defined in [7], Corollary 3.10:*

$$\mathbf{P}[\rho\Delta > x] \to \mathbf{P}[T > x] = \int_0^\infty \frac{e^{-y}}{1 + 2e^{2x}y}\, dy.$$

The errors in these distributional approximations are also quantified, for given choices of L and $\rho(L)$.

This result shows that, for ρ small and $x = l\rho$ with $l \in \mathbf{Z}$,

$$\begin{aligned} \mathbf{P}[\rho\Delta > x] &= \mathbf{E}\{e^{-2\phi_0^2(1+2\rho)^{x/\rho} W_\rho W'_\rho}\} \\ &\approx \mathbf{E}\{e^{-2e^{2x}WW'}\} = \mathbf{P}[T > x], \end{aligned} \tag{2.4}$$

where W and W' are independent NE(1) random variables. Indeed, it follows from Lemma 3.13 below that $W_\rho \to_{\mathcal{D}} W$ as $\rho \to 0$. One way of realizing a random variable T with the above distribution is to realize W and W', and then to sample T from the conditional distribution

$$\begin{aligned}\mathbf{P}[T > x \,|\, W, W'] &= e^{-2e^{2x}WW'} \\ &= e^{-\exp\{2x+\log 2+\log W+\log W'\}} \\ &= e^{-\exp\{2x+\log 2-G_1-G_2\}}, \end{aligned} \tag{2.5}$$

where $G_1 := -\log W$ and $G_2 := -\log W'$ both have the Gumbel distribution. With this construction,

$$\mathbf{P}[2T - \{G_1 + G_2 - \log 2\} > x \,|\, W, W'] = e^{-e^x},$$

whatever the values of W and W', and hence of G_1 and G_2, implying that

$$2T =_{\mathcal{D}} G_1 + G_2 - G_3 - \log 2,$$

where G_1, G_2 and G_3 are independent random variables with the Gumbel distribution. The cumulants of T can thus immediately be deduced from those of the Gumbel distribution, given in Gumbel [9]:

$$\begin{aligned}\mathbf{E}T &= \frac{1}{2}(\gamma - \log 2) \approx -0.058; \\ \operatorname{Var} T &= \frac{\pi^2}{8}.\end{aligned}$$

Note that, in view of Corollary 3.2 below, the conditional construction (2.5) can be interpreted in terms of the processes S and S', since W_ρ and W'_ρ are essentially determined by the early stages of the respective pure birth processes, and the extra randomness, conditional on the values of W_ρ and W'_ρ, comes from the random arrangement of the intervals on the circle C.

In the NMW heuristic, the random variable T_{NMW} is logistic, having distribution function $e^{2x}(1 + e^{2x})^{-1}$; note that this is just the distribution of $\frac{1}{2}(G_1 - G_3)$. Hence the heuristic effectively neglects some of the initial branching variation.

3 The continuous circle model: proofs

The first step in the argument outlined above is to establish a Poisson approximation theorem for the number of pairs of overlapping intervals, one in S_r and the other in $S'_{r'}$. The following result has been shown in [7].

Proposition 3.1 *Let M intervals $I_1, \ldots, I_M$ with lengths $t_1, \ldots, t_M$ and N intervals $J_1, \ldots, J_N$ with lengths $u_1, \ldots, u_N$ be positioned uniformly and independently on C. Set $V := \sum_{i=1}^M \sum_{j=1}^N X_{ij}$, where $X_{ij} := I[I_i \cap J_j \neq \emptyset]$. Then*

$$d_{TV}(\mathcal{L}(V), \mathrm{Po}\,(\lambda_{(M,N,t,u)})) \leq 4(M+N)v_{tu}/L,$$

where $\lambda_{(M,N,t,u)} := L^{-1}(Nt + Mu)$, $t := \sum_{i=1}^M t_i$, $u := \sum_{j=1}^N u_j$ and $v_{tu} := \max\{\max_i t_i, \max_j u_j\}$.

The proposition translates immediately into a useful statement about $\widehat{V}_{r,r'}$, when P' is chosen uniformly at random, independently of all else.

Corollary 3.2 *For the processes S and S' of the previous section, we have*

$$|\mathbf{P}[\widehat{V}_{r,r'} = 0 \,|\, M_r = M, N_{r'} = N, s_r = t, u_{r'} = u] - \exp\{-L^{-1}(Nt+Mu)\}|$$
$$\leq 8L^{-1}(M\tau_r + N\tau_{r'}).$$

Remark. If P' is not chosen at random, but is a fixed point of C, the result of Corollary 3.2 remains essentially unchanged, provided that P and P' are more than an arc distance of $\tau_r + \tau_{r'}$ apart. The only difference is that then $X_{11} = 0$ a.s., and that $Nt + Mu$ is replaced by $Nt + Mu - 2\tau_r - 2\tau_{r'}$. If P and P' are less than $\tau_r + \tau_{r'}$ apart, then $\mathbf{P}[\widehat{V}_{r,r'} = 0] = 0$.

The next step is to show that $\mathbf{P}[\widehat{V}_{r,r'} = 0]$ is close to $\mathbf{P}[V_{r,r'} = 0]$. We do this by directly comparing the random variables $\widehat{V}_{r,r'}$ and $V_{r,r'}$ in the joint construction. As for Corollary 3.5 in [7], the following assertion can easily be shown to hold.

Proposition 3.3 *With notation as above, we have*

$$\mathbf{P}[\widehat{V}_{r,r'} \neq V_{r,r'}] \leq 32\tau_r\tau_{r'}L^{-2}\mathbf{E}\{\tfrac{1}{2}M_r N_{r'}(M_r + N_{r'} - 2)\}.$$

To apply Corollary 3.2 and Proposition 3.3, it remains to establish more detailed information about the distributions of M_r and s_r. In particular, we need to bound the first and second moments of M_r, and to approximate the quantity $\mathbf{E}(\exp\{-L^{-1}(N_{r'}s_r + M_r u_{r'})\})$. We begin with the following lemma.

Lemma 3.4 *The random variable $M(n)$ has as probability generating function*

$$G_{M(n)}(s) := \mathbf{E}s^{M(n)} = f^{(n)}(s), \quad f(s) = se^{2\rho(s-1)},$$

where $f^{(n)}$ denotes the nth iteration of f. In particular, we have

$$\begin{aligned}
\mathbf{E}M_r &= \phi_0\sqrt{L\rho}(1+2\rho)^r \\
\frac{1}{2}\mathbf{E}M_r(M_r-1) &= (\rho+1)\phi_0\sqrt{L\rho}(1+2\rho)^{r-1}\left\{\phi_0\sqrt{L\rho}(1+2\rho)^r - 1\right\} \\
&\leq \phi_0^2 L\rho(1+2\rho)^{2r}.
\end{aligned}$$

Proof: Since $M(n)$ is a branching process with $1 + \mathrm{Po}\,(2\rho)$ offspring distribution, the probability generating function is immediate, as are the moment calculations

$$\begin{aligned}
\mathbf{E}M(n) &= (1+2\rho)^n; \\
\mathbf{E}M(n)(M(n)-1) &= 2(\rho+1)(1+2\rho)^{n-1}\left\{(1+2\rho)^n - 1\right\}.
\end{aligned}$$

The moments of M_r follow from the definition of τ_r. []

These estimates can be directly applied in Corollary 3.2 and Proposition 3.3. Define

$$\eta_1(r,r') := 64\{\rho(n_0 + (r\vee r'))\}^2\phi_0^3(1+2\rho)^{r+r'+(r\vee r')} \tag{3.1}$$

$$\eta_2(r,r') := 16\{\rho(n_0 + (r\vee r'))\}\phi_0(1+2\rho)^{(r\vee r')}. \tag{3.2}$$

Corollary 3.5 *We have*

$$\mathbf{P}[\widehat{V}_{r,r'} \neq V_{r,r'}] \leq \eta_1(r,r')(L\rho)^{-1/2}$$

and

$$|\mathbf{P}[V_{r,r'} = 0] - \mathbf{E}\exp\{-L^{-1}(N_{r'}s_r + M_r u_{r'})\}| \leq \{\eta_1(r,r') + \eta_2(r,r')\}(L\rho)^{-1/2}.$$

Consideration of the quantity $\mathbf{E}(\exp\{-L^{-1}(N_{r'}s_r + M_r u_{r'})\})$ now gives the immediate asymptotics of

$$\mathbf{P}[V_{r,r'} = 0] = \mathbf{P}[D > 2n_0 + r + r'],$$

where D denotes the "small world" distance between P and P'.

Corollary 3.6 *If $\rho = \rho(L)$ is bounded above and $L\rho \to \infty$, then as $L \to \infty$,*

$$|\mathbf{P}[D > 2n_0 + r + r'] - \mathbf{E}\exp\{-2\phi_0^2(1+2\rho)^{r+r'}W_\rho W'_\rho\}| \to 0$$

uniformly in $|r|, |r|' \leq \frac{1}{6\log(1+2\rho)}\log(L\rho)$, where W_ρ and W'_ρ are independent copies of the limiting random variable associated with the pure birth chain M.

Proof: The conditions ensure that τ_r and $\tau_{r'}$ both tend to infinity as $L \to \infty$, at least as fast at $c\log(L\rho)$, for some $c > 0$. Then, since $W(n) = (1+2\rho)^{-n}M(n) \to W_\rho$ a.s. and $s(n)/M(n) \to \rho^{-1}$ a.s., and since $(1+2\rho)^{\tau_r+\tau_{r'}} = \phi_0^2(1+2\rho)^{r+r'}L\rho$, it is clear that

$$\begin{aligned}
\exp\{-L^{-1}(N_{r'}s_r + M_r u_{r'})\} &\sim \exp\{-2(L\rho)^{-1}M_r N_{r'}\} \\
&= \exp\left\{-2(L\rho)^{-1}(1+2\rho)^{\tau_r+\tau_{r'}}W(\tau_r)W'(\tau_{r'})\right\} \\
&\sim \exp\{-2\phi_0^2(1+2\rho)^{r+r'}W_\rho W'_\rho\},
\end{aligned}$$

uniformly for r, r' in the given ranges. []

Hence $\mathbf{P}[D > 2n_0 + r + r']$ can be approximated in terms of the distribution of the limiting random variable W_ρ associated with the pure birth chain M. However, in contrast to the model with time running continuously, this distribution is not always NE(1), but genuinely depends on ρ. Its properties are not so easy to derive, though moments can be calculated, and, in particular,

$$\mathbf{E}W_\rho = 1; \qquad \mathrm{Var}\, W_\rho = 1/(1+2\rho); \tag{3.3}$$

it is also shown in Lemma 3.13 that $\mathcal{L}(W_\rho)$ is close to NE(1) for ρ small. We also need the following lemma, which is useful in bounding the behaviour of the upper tail of $\mathcal{L}(D)$.

Lemma 3.7 *For all $\theta, \rho > 0$,*

$$\mathbf{E}(e^{-\theta W_\rho W'_\rho}) \le \theta^{-1} \log(1+\theta).$$

Proof: The offspring generating function of the birth process M satisfies

$$f(s) = se^{2\rho(s-1)} \le s\{1 + 2\rho(1-s)\}^{-1} =: f_1(s)$$

for all $0 \le s \le 1$. Hence, with $m = 1 + 2\rho$,

$$\mathbf{E}(e^{-\psi W_\rho}) = \lim_{n\to\infty} f^{(n)}(e^{-\psi m^{-n}}) \le \lim_{n\to\infty} f_1^{(n)}(e^{-\psi m^{-n}}) = (1+\psi)^{-1}. \quad (3.4)$$

The last equality follows from (8.11), p.17 in [10], noting that the right-hand side is the Laplace transform of the NE(1) - distribution. Furthermore, we have

$$(1+\theta w)^{-1} = \theta^{-1} \int_0^\infty e^{-tw} e^{-t/\theta}\, dt,$$

and so, applying (3.4) twice, and because the function $(1+t)^{-1}$ is decreasing in $t \ge 0$, we obtain

$$\begin{aligned} \mathbf{E}\left(e^{-\theta W_\rho W'_\rho}\right) &\le \mathbf{E}\{(1+\theta W_\rho)^{-1}\} \\ &= \theta^{-1} \int_0^\infty \mathbf{E} e^{-tW_\rho} e^{-t/\theta}\, dt \\ &\le \theta^{-1} \int_0^\infty (1+t)^{-1} e^{-t/\theta}\, dt \\ &\le \theta^{-1} \int_0^\theta (1+t)^{-1}\, dt = \theta^{-1} \log(1+\theta), \end{aligned}$$

as required. []

The simple asymptotics of Corollary 3.6 can be sharpened. At first sight surprisingly, it turns out that it is not necessary for the times τ_r and $\tau_{r'}$ to tend to infinity, since, for values of ρ so large that n_0 is bounded, the quantities $W(n)$ are (almost) constant for all n. Write

$$\begin{aligned} (1+2\rho)^{-n} s(n) &= 2 \sum_{j=0}^{n-1} W(j)(1+2\rho)^{-(n-j)} \\ &= \frac{1}{\rho} W(n) + \frac{1+\rho}{\rho} U(n), \end{aligned} \quad (3.5)$$

where

$$\frac{1+\rho}{\rho}U(n) \quad = \quad 2\sum_{j=0}^{n-1}(W(j)-W(n))(1+2\rho)^{-(n-j)} - W(n)\rho^{-1}(1+2\rho)^{-n}.$$

Computation gives $\mathbf{E}U(n) = -(1+\rho)^{-1}(1+2\rho)^{-n}$, and

$$\mathbf{E}\{(W(n)-W(j))(W(n)-W(\ell))\} = \frac{1}{(1+2\rho)^{j+1}}\left(1-\frac{1}{(1+2\rho)^{n-j}}\right)$$

if $j \geq \ell$, so that

$$\begin{aligned} \mathrm{Var}\,\{U(n)\} &\leq 2\,\frac{2(1+2\rho)^{-n}+2(1+2\rho)^{-2n}}{(1+\rho)^2} \\ &\leq 8\,\frac{(1+2\rho)^{-n}}{(1+\rho)^2}, \end{aligned} \tag{3.6}$$

and thus

$$(1+\rho)^2\mathbf{E}\{U(n)^2\} \leq 9(1+2\rho)^{-n}. \tag{3.7}$$

Then we have

$$\begin{aligned} &L^{-1}(N_{r'}s_r + M_r u_{r'}) \\ &= \phi_0^2(1+2\rho)^{r+r'}\{W(\tau_r)(W'(\tau_{r'})+(1+\rho)U'_{r'}) + W'(\tau_{r'})(W(\tau_r)+(1+\rho)U_r)\}, \end{aligned}$$

where $W(\tau_r) := W(\tau_r)$ and $U_r := U(\tau_r)$, so that, by Taylor's expansion, and because $\mathbf{E}W(n) = 1$ for all n,

$$\begin{aligned} &\left|\mathbf{E}\exp\{-L^{-1}(N_{r'}s_r + M_r u_{r'})\} - \mathbf{E}\exp\{-2\phi_0^2(1+2\rho)^{r+r'}W(\tau_r)W'(\tau_{r'})\}\right| \\ &\quad \leq \phi_0^2(1+2\rho)^{r+r'}(1+\rho)\{\mathbf{E}|U_r| + \mathbf{E}|U'_{r'}|\} \end{aligned} \tag{3.8}$$

and

$$\begin{aligned} &\left|\mathbf{E}\exp\{-2\phi_0^2(1+2\rho)^{r+r'}W(\tau_r)W'(\tau_{r'})\} - \mathbf{E}\exp\{-2\phi_0^2(1+2\rho)^{r+r'}W_\rho W'_\rho\}\right| \\ &\quad \leq 2\phi_0^2(1+2\rho)^{r+r'}\{\mathbf{E}|W_\rho - W(\tau_r)| + \mathbf{E}|W'_\rho - W'(\tau_{r'})|\}. \end{aligned} \tag{3.9}$$

Using these results, we obtain the following theorem.

Theorem 3.8 *If P' is randomly chosen on C, then*

$$\begin{aligned} &\left|\mathbf{P}[D > 2n_0 + r + r'] - \mathbf{E}\{e^{-2\phi_0^2(1+2\rho)^{r+r'}W_\rho W'_\rho}\}\right| \\ &\quad \leq \{\eta_1(r,r') + \eta_2(r,r')\}(L\rho)^{-1/2} + \eta_3(r,r')(L\rho)^{-1/4}, \end{aligned}$$

where η_1, η_2 are given in (3.1) and (3.2),

$$\eta_3(r,r') := 10\phi_0^{3/2}(1+2\rho)^{r+r'-\frac{1}{2}(r\wedge r')}$$

and where, as before, D denotes the shortest distance between P and P' on the shortcut graph.

Proof: Since $\{V_{r,r'} = 0\} = \{D > 2n_0 + r + r'\}$, we use Corollary 3.5 and (3.8) and (3.9) to give

$$\begin{aligned} &\left|\mathbf{P}[D > 2n_0 + r + r'] - \mathbf{E}\{e^{-2\phi_0^2(1+2\rho)^{r+r'}W_\rho W_\rho{}'}\}\right| \\ &\leq \quad \{\eta_1(r,r') + \eta_2(r,r')\}(L\rho)^{-1/2} + \phi_0^2(1+2\rho)^{r+r'}(1+\rho)\{\mathbf{E}|U_r| + \mathbf{E}|U'_{r'}|\} \\ &\qquad + 2\phi_0^2(1+2\rho)^{r+r'}\{\mathbf{E}|W_\rho - W(\tau_r)| + \mathbf{E}|W'_\rho - W'(\tau_{r'})|\}. \qquad (3.10) \end{aligned}$$

Now, from (3.7) and the Cauchy-Schwarz inequality,

$$(1+\rho)\mathbf{E}|U_r| \leq 3(L\rho)^{-1/4}\phi_0^{-1/2}(1+2\rho)^{-r/2}. \qquad (3.11)$$

Then, since $W(n)$ is a martingale, and

$$\begin{aligned} W_\rho - W(n) &= \sum_{\ell=n}^{\infty}(W(\ell+1) - W(\ell)) \\ &= \sum_{\ell=n}^{\infty}(1+2\rho)^{-\ell-1}(M(\ell+1) - (1+2\rho)M(\ell)), \end{aligned}$$

we have

$$\mathbf{E}(W_\rho - W(n))^2 = \sum_{\ell=n}^{\infty}(1+2\rho)^{-2(\ell+1)}\mathbf{E}(M(\ell+1) - (1+2\rho)M(\ell))^2.$$

Now

$$M(\ell+1) = \sum_{i=1}^{M(\ell)}(Z_{\ell+1}(i) + 1),$$

where $(Z_\ell(i))_{\ell,i}$ are i.i.d. Po (2ρ)-variates, and so

$$\begin{aligned} \mathbf{E}(M(\ell+1) - (1+2\rho)M(\ell))^2 &= \mathbf{E}\mathrm{Var}\,(M(\ell+1)\,|\,M(\ell)) \\ &= 2\rho\mathbf{E}M(\ell) = 2\rho(1+2\rho)^\ell, \end{aligned}$$

implying that

$$\begin{aligned}
\mathbf{E}(W_\rho - W(\tau_r))^2 &\le 2\rho(1+2\rho)^{-2}\sum_{\ell=\tau_r}^{\infty}(1+2\rho)^{-\ell}\\
&= (1+2\rho)^{-2}(1+2\rho)^{-\tau_r}.
\end{aligned}$$

Hence

$$\begin{aligned}
\mathbf{E}|W_\rho - W(\tau_r)| &\le (1+2\rho)^{-1-\frac{1}{2}(n_0+r)}\\
&\le (L\rho)^{-1/4}\phi_0^{-1/2}(1+2\rho)^{-\frac{r}{2}},
\end{aligned}$$

and the theorem follows. []

Theorem 3.8 can be translated into a uniform distributional approximation, as follows.

Theorem 3.9 *If Δ denotes a random variable on the integers with distribution given by*

$$\mathbf{P}\,[\Delta > x] = \mathbf{E}\{e^{-2\phi_0^2(1+2\rho)^x W_\rho W_\rho{}'}\},\quad x \in \mathbf{Z}, \tag{3.12}$$

and $D^* = \Delta + 2n_0$, then

$$\begin{aligned}
&d_{TV}(\mathcal{L}(D), \mathcal{L}(D^*))\\
&= O\left(\log(L\rho)(1+2\rho)^{1/4}(L\rho)^{-\frac{1}{7}} + \left(\frac{\rho\log(L\rho)}{\log(1+2\rho)}\right)^2(1+2\rho)^{1/2}(L\rho)^{-\frac{2}{7}}\right).
\end{aligned}$$

In particular, for $\rho = \rho(L) = O(L^\beta)$ with $\beta < 4/31$,

$$d_{TV}(\mathcal{L}(D), \mathcal{L}(D^*)) \to 0 \quad \text{as} \quad L \to \infty.$$

Proof: It is easy to see that Δ, defined as above, is indeed a random variable. Its upper tail is bounded by Lemma 3.7, which implies that

$$\mathbf{P}[\Delta > x] \le \frac{1}{2\phi_0^2}(1+2\rho)^{-x}(2 + x\log(1+2\rho)) \tag{3.13}$$

for any $x > 0$, since $\phi_0 \le 1$ and $\log(1+2y) \le 2 + \log y$ in $y \ge 1$. Then, for any $x \in \mathbf{Z}$, writing $r(x) = \lfloor x/2 \rfloor$ and $r'(x) = x - r(x) \le (x+1)/2$, it follows from Theorem 3.8 that

$$
\begin{aligned}
&|\mathbf{P}[D > 2n_0 + x] - \mathbf{P}[D^* > 2n_0 + x]| \\
&\quad \le \{\eta_1(r(x), r'(x)) + \eta_2(r(x), r'(x))\}(L\rho)^{-1/2} + \eta_3(r(x), r'(x))(L\rho)^{-1/4} \\
&\quad = O\left(\left(\frac{\rho \log(L\rho)}{\log(1+2\rho)} \right)^2 (1+2\rho)^{\frac{1}{2}} (L\rho)^{-\frac{2}{7}} \right. \\
&\qquad\qquad \left. + \left(\frac{\rho \log(L\rho)}{\log(1+2\rho)} \right) (1+2\rho)^{\frac{1}{2}} (L\rho)^{-\frac{3}{7}} + (1+2\rho)^{\frac{1}{4}} (L\rho)^{-\frac{1}{7}} \right),
\end{aligned}
$$

so long as $x \le \lfloor \frac{\frac{1}{7}\log(L\rho) - 2\log\phi_0}{\log(1+2\rho)} \rfloor$. This is combined with (3.13) evaluated at $x = \lceil \frac{\frac{1}{7}\log(L\rho) - 2\log\phi_0}{\log(1+2\rho)} \rceil$, which gives rise to a term of order $O\Big((L\rho)^{-\frac{1}{7}}\log(L\rho)\Big)$, and the main estimate follows.

The above bound tends to zero as $L \to \infty$ as long as $\rho = \rho(L) = O(L^\beta)$ for $\beta < 4/31$. Thus the theorem is proved. []

For larger ρ and for L large, it is easy to check that n_0 can be no larger than 4, so that interpoint distances are extremely short, few steps in each branching process are needed, and the closeness of $\mathcal{L}(D)$ and $\mathcal{L}(D^*)$ could be justified by direct arguments. Even in the range covered by Theorem 3.9, it is clear that $\mathcal{L}(D)$ becomes concentrated on very few values, once ρ is large, since the factor $2\phi_0^2(1+2\rho)^x$ in the exponent in (3.12) is multiplied by the large factor $(1+2\rho)$ if x is increased by 1. The following corollary makes this more precise.

Corollary 3.10 *If N_0 is such that*

$$(1+2\rho)^{N_0} \le L\rho < (1+2\rho)^{N_0+1},$$

and if $L\rho = (1+2\rho)^{N_0+\alpha}$, for some $\alpha \in [0,1)$, then, taking $x_0 = N_0 - 2n_0 + 1$, we have

$$\mathbf{P}[\Delta \ge x_0] \ge 1 - 2(1+2\rho)^{-\alpha},$$

and

$$\mathbf{P}[\Delta \ge x_0 + 1] = \mathbf{E}\exp\{-2(1+2\rho)^{1-\alpha} W_\rho W'_\rho\} \le \tfrac{1}{2}(1+2\rho)^{-1+\alpha}\log(3+4\rho).$$

Proof: The result follows immediately from Jensen's inequality;

$$\begin{aligned}\mathbf{E}\exp\{-2(1+2\rho)^{-\alpha}W_\rho W'_\rho\} &\geq \exp\left\{-2(1+2\rho)^{-\alpha}\mathbf{E}W_\rho \mathbf{E}W'_\rho\right\} \\ &\geq 1-2(1+2\rho)^{-\alpha}\end{aligned}$$

as $\mathbf{E}W_\rho = 1$, and from Lemma 3.7 with $\theta = 2(1+2\rho)^{1-\alpha}$. []

Thus the distribution is essentially concentrated on the single value x_0 if ρ is large and α is bounded away from 0 and 1. If, for instance, α is close to 1, then both x_0 and x_0+1 may carry appreciable probability.

If $\rho \to \rho_0$ as $L \to \infty$, then the distribution of Δ becomes spread out over $\mathbf{Z}$, converging to a non–trivial limit as $L \to \infty$ along any subsequence such that $\phi_0(L,\rho)$ converges. Both this behaviour and that for larger ρ are quite different from the behaviour found in the continuous model of [7]. However, if ρ becomes smaller, the differences become less; we now show that, as $\rho \to 0$, the distribution of $\rho\Delta$ approaches the limiting distribution of T obtained in [7].

The argument is based on showing that the distribution of W_ρ is close to NE (1). To do so, we employ the characterizing Poincaré equation for Galton–Watson branching processes (see Harris [10], Theorem 8.2, p.15); if

$$\phi_\rho(\theta) = \mathbf{E}e^{-\theta W_\rho}$$

is the Laplace transform of $\mathcal{L}(W_\rho)$, then

$$\phi_\rho((1+2\rho)\theta) = f(\phi_\rho(\theta)). \tag{3.14}$$

We show that when $\rho \approx 0$ then $\phi_\rho(\theta)$ is close to $\phi_e(\theta) = (1+\theta)^{-1}$, the Laplace transform of the NE (1) distribution.

Let

$$\mathcal{G} = \left\{ g : [0,\infty) \to \mathbf{R} :\ \|g\|_{\mathcal{G}} := \sup_{\theta>0} \theta^{-2}|g(\theta)| < \infty \right\},$$

and let

$$\mathcal{H} = \{\chi : [0,\infty) \to \mathbf{R} : \chi(\theta) = 1-\theta+g(\theta) \text{ for some } g \in \mathcal{G}\}.$$

Then $\mathcal{H}$ contains all Laplace transforms of probability distributions with mean 1 and finite variance. On $\mathcal{H}$, define the operator Ψ by

$$(\Psi\chi)(\theta) = f\left(\chi\left(\frac{\theta}{m}\right)\right),$$

where

$$f(s) = se^{2\rho(s-1)}$$

is the probability generating function of $1 + \mathrm{Po}\,(2\rho)$, and $m = 1 + 2\rho > 1$. Thus the Laplace transform ϕ_ρ of interest to us is a fixed point of Ψ.

Lemma 3.11 *The operator Ψ is a contraction, and, for all $\chi, \psi \in \mathcal{H}$,*

$$\|\Psi\chi - \Psi\psi\|_{\mathcal{G}} \quad \le \quad \frac{1}{m}\,\|\chi - \psi\|_{\mathcal{G}}.$$

PROOF: For all $\chi, \psi \in \mathcal{H}$ and $\theta > 0$, we have

$$\begin{aligned}
\theta^{-2}|\Psi\chi(\theta) - \Psi\psi(\theta)| &= \theta^{-2}\left|f\left(\chi\left(\frac{\theta}{m}\right)\right) - f\left(\psi\left(\frac{\theta}{m}\right)\right)\right| \\
&\le \|f\|_\infty\,\theta^{-2}\left|\chi\left(\frac{\theta}{m}\right) - \psi\left(\frac{\theta}{m}\right)\right| \\
&= \theta^{-2}m\left|\chi\left(\frac{\theta}{m}\right) - \psi\left(\frac{\theta}{m}\right)\right| \\
&= m^{-1}(\theta/m)^{-2}\left|\chi\left(\frac{\theta}{m}\right) - \psi\left(\frac{\theta}{m}\right)\right| \\
&\le m^{-1}\|\chi - \psi\|_{\mathcal{G}},
\end{aligned}$$

as required. []

Lemma 3.12 *For the Laplace transform ϕ_e, we have*

$$\|\Psi\phi_e - \phi_e\|_{\mathcal{G}} \le \frac{2\rho^2}{(1+2\rho)^2}.$$

PROOF: For all $\theta > 0$, we have

$$\begin{aligned}
\left|\frac{|\Psi\phi_e(\theta) - \phi_e(\theta)|}{\theta^2}\right| &= \frac{1}{1+\theta}\frac{1}{\theta^2}\left|\left(1 + \frac{2\rho\theta}{m+\theta}\right)e^{-2\frac{\rho\theta}{m+\theta}} - 1\right| \\
&\le \frac{1}{2(1+\theta)\theta^2}\left(\frac{2\rho\theta}{m+\theta}\right)^2,
\end{aligned}$$

using the inequality $|(1+x)e^{-x} - 1| \leq \frac{x^2}{2}$ for $x > 0$. The lemma now follows because $m + \theta > m = 1 + 2\rho$ and $1 + \theta > 1$. $\Box$

Lemmas 3.11 and 3.12 together yield the following result.

Lemma 3.13 *For any* $\rho > 0$,

$$\|\phi_\rho - \phi_e\|_{\mathcal{G}} \quad \leq \quad \frac{\rho}{1+2\rho}.$$

PROOF: With Lemmas 3.11 and 3.12, it follows that

$$\begin{aligned}
\|\phi_\rho - \phi_e\|_{\mathcal{G}} &= \|\Psi\phi_\rho - \phi_e\|_{\mathcal{G}} \\
&\leq \|\Psi\phi_\rho - \Psi\phi_e\|_{\mathcal{G}} + \|\Psi\phi_e - \phi_e\|_{\mathcal{G}} \\
&\leq \frac{1}{m}\|\phi_\rho - \phi_e\|_{\mathcal{G}} + \frac{2\rho^2}{(1+2\rho)^2}.
\end{aligned}$$

Note that indeed $\phi_\rho - \phi_e \in \mathcal{G}$. Thus, since $m > 1$, it follows that

$$\|\phi_\rho - \phi_e\|_{\mathcal{G}} \quad \leq \quad \frac{m}{m-1}\frac{2\rho^2}{(1+2\rho)^2} = \frac{\rho}{1+2\rho},$$

as required. $\Box$

As an immediate consequence, $\mathcal{L}(W_\rho) \to \mathrm{NE}(1)$ as $\rho \to 0$. Theorem 3.14 reformulates this convergence as a pointwise comparison theorem directly relevant to the distribution functions of Δ and T.

Theorem 3.14 *Let* W, W' *be independent* NE (1) *random variables. Then, for all* $\theta > 0$*, we have*

$$\left|\mathbf{E}e^{-\theta W_\rho W_\rho{}'} - \mathbf{E}e^{-\theta W W'}\right| \quad \leq \quad \frac{4\rho}{1+\rho}\,\theta^2.$$

PROOF: We have

$$\begin{aligned}
&\mathbf{E}e^{-\theta W_\rho W_\rho{}'} - \mathbf{E}e^{-\theta W W'} \\
&\quad = \mathbf{E}\{\mathbf{E}(e^{-\theta W_\rho W'_\rho}|W'_\rho)\} - \mathbf{E}\{\mathbf{E}(e^{-\theta W W'}|W')\} \\
&\quad = \mathbf{E}\phi_\rho(\theta W'_\rho) - \mathbf{E}\phi_e(\theta W) \\
&\quad = \mathbf{E}\Psi\phi_\rho(\theta W'_\rho) - \mathbf{E}\Psi\phi_e(\theta W'_\rho) + \mathbf{E}\Psi\phi_e(\theta W'_\rho) - \mathbf{E}\phi_e(\theta W'_\rho) \\
&\qquad\quad + \mathbf{E}\phi_e(\theta W'_\rho) - \mathbf{E}\phi_e(\theta W).
\end{aligned}$$

Since

$$\mathbf{E}\phi_e(\theta W'_\rho) \quad = \quad \mathbf{E}e^{-\theta W W'_\rho} = \mathbf{E}\phi_\rho(\theta W),$$

we obtain from the triangle inequality, (3.3) and Lemmas 3.11, 3.12 and 3.13 that

$$\begin{aligned}
\left|\mathbf{E}e^{-\theta W_\rho W'_\rho} - \mathbf{E}e^{-\theta W W'}\right| &\le \frac{1}{m}\|\phi_\rho - \phi_e\|_{\mathcal{G}}\theta^2\mathbf{E}(W_\rho^2) + \frac{2\rho^2}{(1+2\rho)^2}\theta^2\mathbf{E}(W_\rho^2) \\
&\quad + \|\phi_\rho - \phi_e\|_{\mathcal{G}}\theta^2\mathbf{E}(W_\rho^2) \\
&\le \frac{2\theta^2(1+\rho)}{1+2\rho}\left\{\left(\frac{1}{1+2\rho}+1\right)\frac{\rho}{1+2\rho} + \frac{2\rho^2}{(1+2\rho)^2}\right\} \\
&\le \frac{4\rho}{1+2\rho}\,\theta^2,
\end{aligned}$$

as required. []

Noting that

$$\mathbf{E}e^{-\theta W W'} = \int_0^\infty \frac{e^{-y}}{1+\theta y}\,dy,$$

we obtain the following theorem.

Theorem 3.15 *As in Theorem 3.9, let Δ be a random variable on $\mathbf{Z}$ with distribution given by*

$$\mathbf{P}\left[\Delta > x\right] = \mathbf{E}\{e^{-2\phi_0^2(1+2\rho)^x W_\rho {W_\rho}'}\}.$$

Let T denote a random variable on $\mathbf{R}$ with distribution given by

$$\mathbf{P}\left[T > z\right] = \int_0^\infty \frac{e^{-y}}{1+2ye^{2z}}\,dy.$$

Then

$$\sup_{z\in\mathbf{R}} |\mathbf{P}[\rho\Delta > z] - \mathbf{P}[T > z]| = O\left(\rho^{1/3}(1+\log(1/\rho))\right).$$

PROOF: We use an argument similar to that used for Theorem 3.9. For a large, we can use the bound

$$\int_0^\infty \frac{e^{-y}dy}{1+ay} \le \int_0^1 \frac{dy}{1+ay} = a^{-1}\log(1+a), \tag{3.15}$$

from which, for $z > 0$ and with $c(\rho)$ defined by

$$1 \geq c(\rho) := (2\rho)^{-1} \log(1 + 2\rho) \geq 1 - \rho,$$

we have

$$\mathbf{P}[T > zc(\rho)] \leq e^{-2zc(\rho)}(1 + zc(\rho)) \leq (1 + zc(\rho))e^{-2z(1-\rho)}. \tag{3.16}$$

Similarly, from Lemma 3.7, we have

$$\mathbf{P}[\rho\Delta > z] \leq (1 + 2\rho)^{-(z/\rho)+2}(1 + zc(\rho)) \leq (1 + 2\rho)^2(1 + zc(\rho))e^{-2z(1-\rho)}.$$

Complementing these upper tail bounds, from Theorem 3.14 and for $z \in \rho\mathbf{Z}$, we have

$$\left|\mathbf{P}[\rho\Delta > z] - \int_0^\infty \frac{e^{-y}}{1 + 2y\phi_0^2(1 + 2\rho)^{z/\rho}}\, dy\right| \leq \frac{16\rho}{1 + 2\rho}\,(1 + 2\rho)^{2z/\rho} \leq \frac{16\rho}{1 + 2\rho}\, e^{4z}. \tag{3.17}$$

Using the facts that $(1 + 2\rho)^{z/\rho} = e^{2zc(\rho)}$ and that $(1 + 2\rho)^{-1} \leq \phi_0 \leq 1$, and because, for $a, b > 0$,

$$\left|\int_0^\infty \frac{e^{-y}}{1 + ay}\, dy - \int_0^\infty \frac{e^{-y}}{1 + by}\, dy\right| \leq \frac{|b - a|}{\max\{1, a, b\}}, \tag{3.18}$$

it also follows that

$$\begin{aligned}\left|\int_0^\infty \frac{e^{-y}}{1 + 2y\phi_0^2(1 + 2\rho)^{z/\rho}}\, dy - \mathbf{P}[T_0 > zc(\rho)]\right| \\ \leq 2|\phi_0^2 - 1|(1 + 2\rho)^{z/\rho} \leq \frac{8\rho}{1 + 2\rho}\, e^{2z};\end{aligned} \tag{3.19}$$

and then, from (3.18) and (3.15), we have

$$|\mathbf{P}[T > zc(\rho)] - \mathbf{P}[T > z]| \leq \min\{4z(1 - c(\rho)), (2 + z)e^{-z}\} = O(\rho \log(1/\rho)). \tag{3.20}$$

Combining the bounds (3.17), (3.19) and (3.20) for $e^{2z} \leq \rho^{-1/3}$ gives a supremum of order $\rho^{1/3}$ for $|\mathbf{P}[\rho\Delta > z] - \mathbf{P}[T > z]|$; note that z may actually be allowed to take any real value in this range, since T has bounded density. For any larger values of z, the upper tail bounds give a maximum discrepancy of order $O\{\rho^{1/3}(1 + \log(1/\rho))\}$, as required. Note that, in the main part of the distribution, for z of order 1, the discrepancy is actually of order ρ. []

Numerically, instead of calculating the limiting distribution of W_ρ, we would use the approximation

$$\left|\mathbf{E}\left\{e^{-L^{-1}(N_{r'}s_r+M_r u_{r'})}\right\} - \mathbf{E}\left\{e^{-2\phi_0^2(1+2\rho)^{r+r'}W(\tau_r)W'(\tau_{r'})}\right\}\right| \\ \leq 6\phi_0^{3/2}(1+\rho)(1+2\rho)^{r+r'-\frac{1}{2}(r\wedge r')}(L\rho)^{-1/4},$$

from (3.8) and (3.11), where the distributions of $W(\tau_r)$ and $W'(\tau_{r'})$ can be calculated iteratively, using the generating function from Lemma 3.4. As D is centred around $2n_0 = 2\lfloor \frac{N}{2} \rfloor$, and as r is of order at most $\frac{\log(L\rho)}{\log(1+2\rho)}$, only order $\frac{\log(L\rho)}{\log(1+2\rho)}$ iterations would be needed.

4 The discrete circle model: description

Now suppose, as in the discrete circle model of Newman *et al.* [13], that the circle C becomes a ring lattice with $\Lambda = Lk$ vertices, where each vertex is connected to all its neighbours within distance k by an undirected edge. In the notation of [13], a number of shortcuts are added between randomly chosen pairs of sites, with probability ϕ per connection in the lattice, of which there are Λk; thus, on average, there are $\Lambda k\phi$ shortcuts in the graph. In contrast to the previous setting, it is natural in the discrete model to use graph distance, which implies that all edges, *including shortcuts*, have length 1. This turns out to make a significant difference to the results when shortcuts are very plentiful.

For ease of comparison with the previous model, which collapsed the k-neighbourhoods, we adopt a different notation. The model can be formulated as the union of a Bernoulli random graph $G_{\Lambda,\frac{\sigma}{\Lambda}}$ and the underlying ring lattice on Λ vertices. Here we write $\sigma = \frac{\rho}{k}$, so that the expected number of edges in $G_{\Lambda,\frac{\sigma}{\Lambda}}$ is close to the value $L\rho/2$ in the previous model; comparing the expected number of shortcuts with that given in [13], we also have

$$\Lambda k\phi = \tfrac{1}{2}\Lambda(\Lambda - 2k - 1)\frac{\sigma}{\Lambda} = \frac{\rho}{2k}(\Lambda - 2k - 1) \approx \tfrac{1}{2}L\rho,$$

relating our parameter σ to those of [13]. In particular, we have

$$\sigma = \frac{2\Lambda k\phi}{\Lambda - 2k - 1} \approx 2k\phi. \tag{4.1}$$

The model can also be realized by a dynamic construction. Choosing a point $P_0 \in \{1, \ldots, \Lambda\}$ at random, set $R(0) = \{P_0\}$. Then, at the first step (distance 1), the 'island' consisting of P_0 is increased by k points at each end, and, in addition, $M_1^{(1)} \sim \mathrm{Bi}\,(\Lambda - 2k - 1, \frac{\sigma}{\Lambda})$ shortcuts connect to centres of new islands. At each subsequent step, starting from the set $R(n)$ of vertices within distance n of P_0, each island is increased by the addition of k points at either end, but with overlapping islands merged, to form a set $R'(n+1)$; this is then increased to $R(n+1)$ by choosing a Bernoulli–$\frac{\sigma}{\Lambda}$ thinning of the edges joining $R(n) \setminus R(n-1)$ to $C \setminus R'(n+1)$ as shortcuts.

The branching analogue of this process, which agrees with the current process until its first self–overlap occurs, has individuals, here representing the islands, of two types: newly formed type 1 islands, consisting of just one vertex, and existing type 2 islands. A type 1 island at time n becomes a type 2 island at time $n+1$, and, in addition, has a $\mathrm{Bi}\,(\Lambda, \frac{\sigma}{\Lambda})$–distributed number of type 1 islands as 'offspring'. A type 2 island at time n stays a type 2 island at time $n+1$, and has a $\mathrm{Bi}\,(2k\Lambda, \frac{\sigma}{\Lambda})$–distributed number of type 1 islands as offspring. Each new island starts at an independent and randomly chosen point of the circle, and at each subsequent step acquires k more vertices at either end. Writing

$$\hat{M}(n) := (\hat{M}^{(1)}(n), \hat{M}^{(2)}(n))^T, \quad n \geq 0,$$

for the numbers of islands of the two types at time n, where the superscript T denotes the transpose, their development over time is given by the branching recursion

$$\begin{aligned}
\hat{M}^{(1)}(n) &\sim \mathrm{Bi}\left((\hat{M}^{(1)}(n-1) + 2k\hat{M}^{(2)}(n-1))\Lambda, \frac{\sigma}{\Lambda}\right), \\
\hat{M}^{(2)}(n) &= \hat{M}^{(1)}(n-1) + \hat{M}^{(2)}(n-1) : \\
\hat{M}^{(1)}(0) &= 1, \quad \hat{M}^{(2)}(0) = 0.
\end{aligned} \tag{4.2}$$

The total number of intervals at time n is denoted by

$$\hat{M}^{+}(n) = \hat{M}^{(1)}(n) + \hat{M}^{(2)}(n), \tag{4.3}$$

and the total number of vertices in these intervals by

$$\hat{s}(n) = \hat{M}^+(n) + 2k \sum_{j=0}^{n-1} \hat{M}^+(j) \geq \hat{M}^+(n). \tag{4.4}$$

As before, we use the branching process as the basic tool in our argument. It is now a two type Galton–Watson process with mean matrix

$$A = \begin{pmatrix} \sigma & 2k\sigma \\ 1 & 1 \end{pmatrix}.$$

The characteristic equation

$$(t-1)(t-\sigma) = 2k\sigma \tag{4.5}$$

of A yields the eigenvalues

$$\begin{aligned} \lambda \,=\, \lambda_1 \quad &= \quad \tfrac{1}{2}\{\sigma + 1 + \sqrt{(\sigma+1)^2 + 4\sigma(2k-1)}\} > \sigma + 1; \\ -\lambda \,<\, \lambda_2 \quad &= \quad \tfrac{1}{2}\{\sigma + 1 - \sqrt{(\sigma+1)^2 + 4\sigma(2k-1)}\} < 0 : \end{aligned}$$

also, from (4.5),

$$\lambda + \lambda_2 = \sigma + 1 \quad \text{and} \quad \lambda\lambda_2 = -\sigma(2k-1). \tag{4.6}$$

From the equation $fA = \lambda f$, we find that the left eigenvectors $f^{(i)}$, $i = 1, 2$, satisfy

$$f_2^{(i)} = (\lambda_i - \sigma) f_1^{(i)}. \tag{4.7}$$

We standardize the positive left eigenvector $f^{(1)}$ of A, associated with the eigenvalue λ, so that

$$f_1^{(1)} \;\;=\;\; (\lambda - \sigma)^{-\frac{1}{2}}, \quad f_2^{(1)} \;=\; (\lambda - \sigma)^{\frac{1}{2}}; \tag{4.8}$$

for $f^{(2)}$, we choose

$$f_1^{(2)} = (\sigma - \lambda_2)^{-\frac{1}{2}}, \quad f_2^{(2)} = -(\sigma - \lambda_2)^{\frac{1}{2}}.$$

Then, for $i = 1, 2$, we have

$$\mathbf{E}((f^{(i)})^T \hat{M}_{n+1} | \mathcal{F}(n)) \;\;=\;\; (f^{(i)})^T A \hat{M}(n) = \lambda_i (f^{(i)})^T \hat{M}(n),$$

where $\mathcal{F}(n)$ denotes the σ-algebra $\sigma(\hat{M}(0),\ldots,\hat{M}(n))$. Thus, from (4.7),

$$\begin{aligned} W^{(i)}(n) &:= \lambda_i^{-n}(f^{(i)})^T\hat{M}(n) \\ &= \lambda_i^{-n}f_1^{(i)}(\hat{M}_n^{(1)}+(\lambda_i-\sigma)\hat{M}_n^{(2)}) \end{aligned} \tag{4.9}$$

is a (non-zero mean) martingale, for $i=1,2$; we let

$$W_{k,\sigma} := \lim_{n\to\infty} W^{(1)}(n) \text{ a.s.} = \lim_{n\to\infty} \lambda_1^{-n}(f^{(1)})^T\hat{M}(n) \text{ a.s.} \tag{4.10}$$

be the almost sure limit of the martingale $W^{(1)}(n)$.

Our main conclusions can be summarized as follows: the detailed results and their proofs are given in Theorems 5.6 and 5.9. Let Δ_d denote a random variable on the integers with distribution given by

$$\mathbf{P}\left[\Delta_d > x\right] = \mathbf{E}\exp\left\{-\left(\frac{\lambda}{\lambda-\lambda_2}\right)^2(\lambda-\sigma)(2\lambda-\sigma)\phi_d^2\lambda^x W_{k,\sigma}W'_{k,\sigma}\right\}, \tag{4.11}$$

for any $x\in\mathbf{Z}$, and set $D^* = \Delta_d + 2n_d$. Here, n_d and ϕ_d are such that $\lambda^{n_d} = \phi_d(\Lambda\sigma)^{1/2}$ and $\lambda^{-1} < \phi_d \le 1$. Let D denote the graph distance between a randomly chosen pair of vertices on the ring lattice C.

Theorem 4.1 *If $\Lambda\sigma\to\infty$ and $\rho=k\sigma$ remains bounded, then it follows that $d_{TV}(\mathcal{L}(D),\mathcal{L}(D^*))\to 0$. If $\rho\to 0$, then $\rho\Delta_d \to_{\mathcal{D}} T$, where T is as in Theorem 2.1.*

Note that the expectation in (4.11) is taken under the initial condition (4.2); we shall later need also to consider the distribution of $W_{k,\sigma}$ under other initial conditions.

5 The discrete circle model: proofs

We begin the detailed discussion with some moment formulae.

Lemma 5.1 *For the means,*

$$\mathbf{E}\hat{M}^{(1)}(n) = \frac{1}{\lambda-\lambda_2}(\lambda^n(\sigma-\lambda_2)+\lambda_2^n(\lambda-\sigma));$$

$$
\begin{aligned}
\mathbf{E}\hat{M}^{(2)}(n) &= \frac{1}{\lambda - \lambda_2}(\lambda^n - \lambda_2^n); \\
\mathbf{E}\hat{M}^{+}(n) &= \frac{1}{\lambda - \lambda_2}(\lambda^{n+1} - \lambda_2^{n+1}) \leq 2\lambda^n, \qquad (5.1)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{E}(\hat{M}_n^{(1)} + 2k\hat{M}_n^{(2)}) &= \frac{1}{(\lambda - \lambda_2)}\{(1 - \lambda_2/\sigma)\lambda^{n+1} + (\lambda_2/\sigma)(\lambda - \sigma)\lambda_2^n\} \\
&\leq c\lambda^n, \ \textit{for } c = 4k - 1. \qquad (5.2)
\end{aligned}
$$

For the variances, for $j \leq n$,

$$
\mathrm{Var}\,(W^{(1)}(j) - W^{(1)}(n)) \leq \kappa^2 (f_1^{(1)})^2 \lambda^{-j}, \qquad \kappa^2 := \frac{c\sigma}{\lambda(\lambda - 1)}; \qquad (5.3)
$$

$$
\mathrm{Var}\,(W^{(2)}(j) - W^{(2)}(n)) \qquad (5.4)
$$

$$
\leq c\sigma \left(\frac{f_1^{(2)}}{\lambda_2}\right)^2 \min\left\{\frac{\lambda_2^2}{|\lambda - \lambda_2^2|}, (n - j)\right\} \left(\frac{\lambda}{\lambda_2^2}\right)^j \max\left\{1, \frac{\lambda}{\lambda_2^2}\right\}^{n-j}
$$

and, for $\hat{M}^+(n)$,

$$
\begin{aligned}
\mathrm{Var}\,\hat{M}^+(n) &\leq 4\kappa^2\lambda^{2n}; \qquad (5.5) \\
\mathbf{E}\hat{M}^+(n)(\hat{M}^+(n) - 1) &\leq 4(1 + \kappa^2)\lambda^{2n}. \qquad (5.6)
\end{aligned}
$$

Note that, from (4.5), we have

$$
0 \leq \kappa^2 = \frac{c}{2k + \lambda - 1} \leq \frac{4k - 1}{2k} \leq 2. \qquad (5.7)
$$

Proof: First, observe that

$$
\mathbf{E}W^{(i)}(n) = W_0^{(i)} = (f^{(i)})^T \hat{M}_0^+ = (f^{(i)})^T (1, 0)^T = f_1^{(i)} \qquad (5.8)
$$

for all n, by the martingale property. From (4.9) and (4.7), we have

$$
\begin{aligned}
(f_1^{(1)})^{-1}\lambda^n W^{(1)}(n) &= \hat{M}^{(1)}(n) + (\lambda - \sigma)\hat{M}^{(2)}(n); \\
(f_1^{(2)})^{-1}\lambda_2^n W^{(2)}(n) &= \hat{M}^{(1)}(n) + (\lambda_2 - \sigma)\hat{M}^{(2)}(n),
\end{aligned}
$$

and thus

$$
\begin{aligned}
\hat{M}^{(1)}(n) &= \lambda^n W^{(1)}(n)\frac{\sigma - \lambda_2}{(\lambda - \lambda_2)f_1^{(1)}} + \lambda_2^n W^{(2)}(n)\frac{\lambda - \sigma}{(\lambda - \lambda_2)f_1^{(2)}}; \qquad (5.9) \\
\hat{M}^{(2)}(n) &= \lambda^n W^{(1)}(n)\frac{1}{(\lambda - \lambda_2)f_1^{(1)}} - \lambda_2^n W^{(2)}(n)\frac{1}{(\lambda - \lambda_2)f_1^{(2)}}.
\end{aligned}
$$

From (5.9) and (5.8) we obtain

$$\begin{aligned}
\mathbf{E}\hat{M}^{(1)}(n) &= \lambda^n \frac{\sigma - \lambda_2}{(\lambda - \lambda_2)} + \lambda_2^n \frac{\lambda - \sigma}{(\lambda - \lambda_2)}; \\
\mathbf{E}\hat{M}^{(2)}(n) &= \lambda^n \frac{1}{(\lambda - \lambda_2)} - \lambda_2^n \frac{1}{(\lambda - \lambda_2)},
\end{aligned}$$

giving (5.1); for the last part use $\sigma + 1 - \lambda = \lambda_2$ and $\sigma + 1 - \lambda_2 = \lambda$, from (4.6). Then (5.2) follows immediately, using (4.5) and (4.6).

Now define

$$X(n) := \hat{M}^{(1)}(n) - \sigma(\hat{M}^{(1)}(n-1) + 2k\hat{M}^{(2)}(n-1)), \quad n \geq 1, \tag{5.10}$$

noting that it has a centred binomial distribution conditional on $\mathcal{F}(n-1)$; representing quantities in terms of these martingale differences greatly simplifies the subsequent calculations. For instance,

$$\begin{aligned}
&W^{(i)}(n+1) - W^{(i)}(n) \\
&\quad = \lambda_i^{-n-1} f_1^{(i)} \{\hat{M}^{(1)}(n+1) + (\lambda_i - \sigma)\hat{M}^{(2)}(n+1) \\
&\qquad\qquad - \lambda_i \hat{M}^{(1)}(n) - \lambda_i(\lambda_i - \sigma)\hat{M}^{(2)}(n)\} \\
&\quad = \lambda_i^{-n-1} f_1^{(i)} \{\hat{M}^{(1)}(n+1) - \sigma\hat{M}^{(1)}(n) - 2k\sigma\hat{M}^{(2)}(n)\} \\
&\quad = \lambda_i^{-n-1} f_1^{(i)} X(n+1),
\end{aligned} \tag{5.11}$$

where we have used $(\lambda_i - 1)(\lambda_i - \sigma) = 2k\sigma$, from (4.5), and the branching recursion.

Since

$$\mathbf{E}\{X^2(n+1) \,|\, \mathcal{F}(n)\} = \frac{\sigma}{\Lambda}\left(1 - \frac{\sigma}{\Lambda}\right)(\hat{M}^{(1)}(n) + 2k\hat{M}^{(2)}(n))\Lambda,$$

we have

$$\mathbf{E}X^2(n+1) \leq c\sigma\lambda^n,$$

from (5.2). Thus, immediately,

$$\mathbf{E}\{(W^{(i)}(n+1) - W^{(i)}(n))^2\} \leq c\sigma(f_1^{(i)})^2 \lambda_i^{-2n-2}\lambda^n. \tag{5.12}$$

Hence, for $i = 1, 2$ and for any $0 \leq j < n$,

$$\mathrm{Var}\,(W^{(i)}(j) - W^{(i)}(n))$$

$$
\begin{aligned}
&= \sum_{k=j}^{n-1} \mathbf{E}(W^{(i)}(k) - W^{(i)}(k+1))^2 \\
&\le c\sigma (f_1^{(i)})^2 \sum_{k=j}^{n-1} \lambda^k \lambda_i^{-2(k+1)} \\
&\le c\sigma (f_1^{(i)})^2 \lambda_i^{-2} \left(\frac{\lambda}{\lambda_i^2}\right)^j \min\left\{\frac{\lambda_i^2}{|\lambda - \lambda_i^2|}, (n-j)\right\} \max\left(1, \left(\frac{\lambda}{\lambda_i^2}\right)^{n-j}\right),
\end{aligned}
$$

and the formulae (5.3) and (5.4) follow.

Moreover, from (5.9),

$$
\hat{M}^+(n) = \frac{1}{\lambda - \lambda_2}\left(\frac{\lambda^{n+1}}{f_1^{(1)}} W^{(1)}(n) - \frac{\lambda_2^{n+1}}{f_1^{(2)}} W^{(2)}(n)\right), \tag{5.13}
$$

and hence

$$
\begin{aligned}
(\lambda - \lambda_2)^2 \operatorname{Var} \hat{M}^+(n) &= \sum_{j=1}^{n} (\lambda^{n+1-j} - \lambda_2^{n+1-j})^2 \operatorname{Var} X(j) \\
&\le 4c\sigma \sum_{j=1}^{n} \lambda^{2n-(j-1)} \le 4c\sigma \lambda^{2n+1}/(\lambda - 1).
\end{aligned}
$$

From this, using the inequality

$$
\mathbf{E}\hat{M}^+(n)(\hat{M}^+(n) - 1) \quad \le \quad \operatorname{Var}(\hat{M}^+(n)) + (\mathbf{E}\hat{M}^+(n))^2,
$$

(5.6) is easily obtained. []

As in the previous section, we run two branching processes $\hat{M}$ and $\hat{M}' =: \hat{N}$ independently, and investigate the time at which the first intersection occurs, irrespective of the types of the intervals. We write $\hat{s}'(n) =: \hat{u}(n)$, and use notation of the form $\hat{M}_r$ to denote $\hat{M}(n_d + r)$, for an appropriate n_d which we shall define later; we also use $\tau_r := \{2k(n_d + r) + 1\}$ to denote the length of the longest interval in the branching process at time $n_d + r$. Then, with $\widehat{V}_{r,r'}$ defined as before to be the number of pairs of intervals of $\hat{M}$ and $\hat{N}$ intersecting, when $\hat{M}$ has been run for time $n_d + r$ and $\hat{N}$ for time $n_d + r'$, the analogue of Proposition 3.3 shows that

$$
\begin{aligned}
&|\mathbf{P}[\widehat{V}_{r,r'} = 0] - \mathbf{P}[D > 2n_d + r + r']| \\
&\qquad \le 32\Lambda^{-2}\tau_{(r\vee r')}^2 \mathbf{E}\{\tfrac{1}{2}\hat{M}_r^+ \hat{N}_{r'}^+ (\hat{M}_r^+ + \hat{N}_{r'}^+ - 2)\},
\end{aligned} \tag{5.14}
$$

and that of Corollary 3.2 gives

$$\begin{aligned}
&|\mathbf{P}[\widehat{V}_{r,r'} = 0 \,|\, \hat{M}_r^+ = M, \hat{N}_{r'}^+ = N, \hat{s}_r = t, \hat{u}_{r'} = u] \\
&\qquad\qquad - \exp\{-\Lambda^{-1}(Nt + Mu - MN)\}| \\
&\quad \le 8\Lambda^{-1}(M+N)\tau_{(r\vee r')}.
\end{aligned} \tag{5.15}$$

The estimates (5.14) and (5.15) can be made more explicit with the help of the bounds

$$\mathbf{E}\hat{M}_r^+ \le 2\lambda^{n_d+r}; \qquad \mathbf{E}\hat{M}_r^+(\hat{M}_r^+ - 1) \le 4(1+\kappa^2)\lambda^{2(n_d+r)}, \tag{5.16}$$

which follow from from Lemma 5.1; together, they give the following result, in which D denotes the shortest distance between P_0 and a randomly chosen vertex P' of C.

Lemma 5.2 *With the above notation and definitions, we have*

$$\begin{aligned}
&|\mathbf{P}[D > 2n_d + r + r'] - \mathbf{P}[\widehat{V}_{r,r'} = 0]| \\
&\qquad \le 256\Lambda^{-2}\tau^2_{(r\vee r')}(1+\kappa^2)\lambda^{3n_d+r+r'+(r\vee r')},
\end{aligned}$$

and

$$\begin{aligned}
&|\mathbf{P}[\widehat{V}_{r,r'} = 0 \,|\, \hat{M}_r^+ = M, \hat{N}_{r'}^+ = N, \hat{s}_r = t, \hat{u}_{r'} = u] \\
&\qquad\qquad - \exp\{-\Lambda^{-1}(Nt + Mu - MN)\}| \\
&\quad \le 32\Lambda^{-1}\tau_{(r\vee r')}\lambda^{n_d+(r\vee r')}.
\end{aligned}$$

We now need to examine $\mathbf{E}\exp\{-\Lambda^{-1}(\hat{N}_{r'}^+\hat{s}_r + \hat{M}_r^+\hat{u}_{r'} - \hat{M}_r^+\hat{N}_{r'}^+)\}$ more closely.

To start with, from (5.1) in Lemma 5.1, we have

$$\begin{aligned}
\mathbf{E}\hat{s}(n) &= \frac{1}{\lambda-\lambda_2}\left\{(\lambda^{n+1} - \lambda_2^{n+1}) + 2k\sum_{j=0}^{n-1}(\lambda^{j+1} - \lambda_2^{j+1})\right\} \\
&= \frac{1}{\lambda-\lambda_2}\left\{\lambda^{n+1}\left(1 + \frac{2k}{\lambda-1}\right) - \lambda_2^{n+1}\left(1 + \frac{2k}{\lambda_2-1}\right)\right. \\
&\qquad\qquad \left. - 2k\left(\frac{\lambda}{\lambda-1} - \frac{\lambda_2}{\lambda_2-1}\right)\right\} \\
&= \frac{1}{\sigma(\lambda-\lambda_2)}\left\{\lambda^{n+2} - \lambda_2^{n+2} - (\lambda-\lambda_2)\right\},
\end{aligned} \tag{5.17}$$

where we have used (4.5) and (4.6) to simplify, and this expression is rather close to $(\lambda/\sigma)\mathbf{E}\hat{M}^+(n)$ as given in (5.1). This reflects the fact that both $\hat{s}(n)$ and $(\lambda/\sigma)\hat{M}^+(n)$ are rather close to

$$\frac{\lambda^{n+2}}{\sigma(\lambda-\lambda_2)}\,\frac{W^{(1)}(n-1)}{f_1^{(1)}}.$$

Lemma 5.3 *We have the following approximations:*

$$\begin{aligned}\hat{s}(n) &= \frac{\lambda^{2+n}}{\sigma(\lambda-\lambda_2)}\left(\frac{W^{(1)}(n-1)}{f_1^{(1)}}+\widetilde{U}_1(n)\right);\\ \frac{\lambda}{\sigma}\hat{M}^+(n) &= \frac{\lambda^{2+n}}{\sigma(\lambda-\lambda_2)}\left(\frac{W^{(1)}(n-1)}{f_1^{(1)}}+\widetilde{U}_2(n)\right),\end{aligned}$$

where

$$\mathbf{E}|\widetilde{U}_1(n)| \leq \{3+2\kappa\sqrt{n+\sigma^2}\}\max\left\{\lambda^{-1/2},\frac{|\lambda_2|}{\lambda}\right\}^n; \tag{5.18}$$

$$\mathbf{E}|\widetilde{U}_2(n)| \leq \{1+2\kappa\sqrt{n+\lambda-1}\}\max\left\{\lambda^{-1/2},\frac{|\lambda_2|}{\lambda}\right\}^n. \tag{5.19}$$

PROOF: We first express $\hat{s}(n)$ and $(\lambda/\sigma)\hat{M}^+(n)$ in terms of the martingale differences $\{X(l),\, l\geq 1\}$. From (5.13) and (5.11), we have

$$\begin{aligned}\hat{M}^+(n) &= \frac{1}{\lambda-\lambda_2}\left\{\frac{\lambda^{n+1}}{f_1^{(1)}}W^{(1)}(n)-\frac{\lambda_2^{n+1}}{f_1^{(2)}}W^{(2)}(n)\right\}\\ &= \frac{\lambda^{n+1}}{(\lambda-\lambda_2)f_1^{(1)}}W^{(1)}(n-1)+X(n)-\frac{\lambda_2^{n+1}}{\lambda-\lambda_2}\left(1+\sum_{l=1}^{n-1}\lambda_2^{-l}X(l)\right).\end{aligned}$$

Similarly, from (5.13) and (5.11),

$$\begin{aligned}&\sum_{j=0}^{n-1}\hat{M}^+(j)\\ &= \frac{1}{\lambda-\lambda_2}\sum_{j=0}^{n-1}\left\{\frac{\lambda^{j+1}}{f_1^{(1)}}W^{(1)}(j)-\frac{\lambda_2^{j+1}}{f_1^{(2)}}W^{(2)}(j)\right\}\\ &= \frac{1}{\lambda-\lambda_2}\sum_{j=0}^{n-1}\left\{\frac{\lambda^{j+1}}{f_1^{(1)}}W^{(1)}(n-1)-\lambda^{j+1}\sum_{l=j+1}^{n-1}\lambda^{-l}X(l)\right.\\ &\qquad\left.-\lambda_2^{j+1}\left(1+\sum_{l=1}^{j}\lambda_2^{-l}X(l)\right)\right\}\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\lambda-\lambda_2}\left\{\frac{W^{(1)}(n-1)}{f_1^{(1)}}\,\frac{\lambda^{n+1}}{\lambda-1}-\frac{\lambda}{\lambda-1}\left(1+\sum_{j=1}^{n-1}\lambda^{-j}X(j)\right)\right.\\
&\qquad\left.-\sum_{l=1}^{n-1}\lambda^{-l}X(l)\sum_{j=0}^{l-1}\lambda^{j+1}-\frac{\lambda_2^{n+1}-\lambda_2}{\lambda_2-1}-\sum_{l=1}^{n-1}\lambda_2^{-l}X(l)\sum_{j=l+1}^{n}\lambda_2^{j}\right\}\\
&= \frac{1}{\lambda-\lambda_2}\left\{\frac{W^{(1)}(n-1)}{f_1^{(1)}}\,\frac{\lambda^{n+1}}{\lambda-1}-\left(\frac{\lambda}{\lambda-1}-\frac{\lambda_2}{\lambda_2-1}\right)\left(1+\sum_{l=1}^{n-1}X(l)\right)\right.\\
&\qquad\left.-\frac{\lambda_2^{n+1}}{\lambda_2-1}\left(1+\sum_{l=1}^{n-1}\lambda_2^{-l}X(l)\right)\right\}.
\end{aligned}
$$

Substituting these into (4.4), and because $1+2k/(\lambda_i-1)=\lambda_i/\sigma,\ i=1,2,$ from (4.5) and (4.6), we obtain

$$
\begin{aligned}
\hat{s}(n) &= \frac{\lambda^{2+n}}{\sigma(\lambda-\lambda_2)}\,\frac{W^{(1)}(n-1)}{f_1^{(1)}}+X(n)\\
&\qquad-\frac{1}{\sigma}\left\{\sum_{l=1}^{n-1}X(l)\left(1+\frac{\lambda_2^{n+2-l}}{\lambda-\lambda_2}\right)+1+\frac{\lambda_2^{n+2}}{\lambda-\lambda_2}\right\}\\
&= \frac{\lambda^{2+n}}{\sigma(\lambda-\lambda_2)}\left(\frac{W^{(1)}(n-1)}{f_1^{(1)}}+\widetilde{U}_1(n)\right),
\end{aligned}
$$

where

$$
|\mathbf{E}\widetilde{U}_1(n)| = \lambda^{-n-2}|\lambda-\lambda_2+\lambda_2^{n+2}| \le 3\max\left\{\lambda^{-1},\frac{|\lambda_2|}{\lambda}\right\}^{n+1}
$$

and

$$
\begin{aligned}
\operatorname{Var}\widetilde{U}_1(n) &\le \lambda^{-2n-4}c\sigma(\lambda-\lambda_2)^2\left\{\sigma^2\lambda^{n-1}+\sum_{l=1}^{n-1}\lambda^{l-1}\left(1+\frac{\lambda_2^{n+2-l}}{\lambda-\lambda_2}\right)^2\right\}\\
&\le 4\kappa^2(\sigma^2+n)\max\left\{\lambda^{-1},\frac{\lambda_2^2}{\lambda^2}\right\}^{n+1},
\end{aligned}
$$

giving the first approximation. By similar arguments, for $(\lambda/\sigma)\hat{M}^{+}(n)$ we obtain

$$
\begin{aligned}
\frac{\lambda}{\sigma}\hat{M}^{+}(n) &= \frac{\lambda^{2+n}}{\sigma(\lambda-\lambda_2)}\,\frac{W^{(1)}(n-1)}{f_1^{(1)}}\\
&\qquad+\frac{\lambda}{\sigma}\left\{X(n)-\frac{1}{\lambda-\lambda_2}\sum_{l=1}^{n-1}X(l)\lambda_2^{n+1-l}-\frac{\lambda_2^{n+1}}{\lambda-\lambda_2}\right\}\\
&= \frac{\lambda^{2+n}}{\sigma(\lambda-\lambda_2)}\left(\frac{W^{(1)}(n-1)}{f_1^{(1)}}+\widetilde{U}_2(n)\right),
\end{aligned}
$$

where

$$|\mathbf{E}\widetilde{U}_2(n)| = (|\lambda_2|/\lambda)^{n+1}$$

and

$$\begin{aligned}\operatorname{Var}\widetilde{U}_2(n) &\le \lambda^{-2n-2}c\sigma(\lambda-\lambda_2)^2\left\{\lambda^{n-1} + \sum_{l=1}^{n-1}\frac{\lambda^{l-1}\lambda_2^{2(n+1-l)}}{(\lambda-\lambda_2)^2}\right\} \\ &\le 4\kappa^2(\lambda-1+n)\max\left\{\lambda^{-1}, \frac{\lambda_2^2}{\lambda^2}\right\}^n,\end{aligned}$$

giving the second approximation. []

We now use these approximations as in the previous section, starting by observing that

$$\begin{aligned}\varepsilon(n,n') := \Bigg|&\mathbf{E}\exp\{-\Lambda^{-1}(\hat{N}^+(n')\hat{s}(n) + \hat{M}^+(n)\hat{u}(n') - \hat{M}^+(n)\hat{N}^+(n'))\} \\ &- \mathbf{E}\exp\left\{-\left(\frac{\lambda}{\lambda-\lambda_2}\right)^2(\lambda-\sigma)(2\lambda-\sigma)\left(\frac{\lambda^{n+n'}}{\Lambda\sigma}\right)W^{(1)}(n-1)W'^{(1)}(n'-1)\right\}\Bigg| \\ \le\ & \frac{\lambda^3}{\sigma(\lambda-\lambda_2)^2}\frac{\lambda^{n+n'}}{\Lambda}\Big\{\mathbf{E}|\widetilde{U}_1'(n')|\mathbf{E}\{\lambda^{-n}\hat{M}^+(n)\} + \mathbf{E}|\widetilde{U}_2(n)| \\ &\qquad + \mathbf{E}|\widetilde{U}_1(n)|\mathbf{E}\{\lambda^{-n'}\hat{N}^+(n')\} + \mathbf{E}|\widetilde{U}_2'(n')|\Big\} \\ &+ \left(\frac{\lambda}{\lambda-\lambda_2}\right)^2\frac{\lambda^{n+n'}}{\Lambda}\Big\{\mathbf{E}|\widetilde{U}_2(n)|\mathbf{E}\{\lambda^{-n'}\hat{N}^+(n')\} + \mathbf{E}|\widetilde{U}_2'(n')|\Big\},\end{aligned}$$

since $\mathbf{E}W^{(1)}(m) = f_1^{(1)} = (\lambda-\sigma)^{-1/2}$ for all m. Since also, from (5.1), $\lambda^{-m}\mathbf{E}\hat{M}^+(m) \le 2$, it follows from Lemma 5.3 and because $\lambda \ge 1+\sigma > 1$ that

$$\begin{aligned}\varepsilon(n,n') &\le \{17 + 18\kappa\sqrt{(n\vee n') + (\sigma^2\vee(\lambda-1))}\}\frac{\lambda^3}{(\lambda-\lambda_2)^2}\frac{\lambda^{n+n'}}{\Lambda\sigma} \\ &\qquad \times \max\left\{\lambda^{-1/2}, \frac{|\lambda_2|}{\lambda}\right\}^{(n\wedge n')}. \qquad (5.20)\end{aligned}$$

Then similarly, since $(W^{(1)}(j) - W_{k,\sigma})/f_1^{(1)}$ has mean zero and, letting $n\to\infty$ in (5.3), variance at most $\kappa^2\lambda^{-j}$, it follows that

$$\left|\mathbf{E}\exp\left\{-\left(\frac{\lambda}{\lambda-\lambda_2}\right)^2(\lambda-\sigma)(2\lambda-\sigma)\left(\frac{\lambda^{n+n'}}{\Lambda\sigma}\right)W^{(1)}(n-1)W'^{(1)}(n'-1)\right\}\right.$$

$$- \mathbf{E} \exp\left\{ -\left(\frac{\lambda}{\lambda - \lambda_2}\right)^2 (\lambda - \sigma)(2\lambda - \sigma) \left(\frac{\lambda^{n+n'}}{\Lambda\sigma}\right) W_{k,\sigma} W'_{k,\sigma} \right\}\Bigg|$$

$$\leq \quad 2\left(\frac{\lambda}{\lambda - \lambda_2}\right)^2 (\lambda - \sigma)(2\lambda - \sigma) \left(\frac{\lambda^{n+n'}}{\Lambda\sigma}\right) \kappa\lambda^{-\{(n\wedge n')-1\}/2}. \qquad (5.21)$$

So choose n_d so that $\lambda^{n_d} = \phi_d(\Lambda\sigma)^{1/2}$ with $\lambda^{-1} < \phi_d \leq 1$, and let $n = n_d + r$, $n' = n_d + r'$; then define the quantities

$$\begin{aligned}
\eta'_1(r, r') &:= 256\phi_d^3\{\sigma(2k(n_d + (r \vee r')) + 1)\}^2(1 + \kappa^2)\lambda^{r+r'+(r\vee r')};\\
\eta'_2(r, r') &:= 32\phi_d\{\sigma(2k(n_d + (r \vee r')) + 1)\}\lambda^{(r\vee r')};\\
\eta'_3(r, r') &:= 4\phi_d^{3/2}\lambda^{1/2}(\lambda - \sigma)^2\kappa\lambda^{r+r'-(r\wedge r')/2},
\end{aligned}$$

and

$$\eta'_4(r, r') := 18\phi_d^{2-\gamma}\lambda^{1-\gamma}\{1 + \kappa(n_d + (r \vee r') + (\sigma^2 \vee (\lambda - 1)))^{1/2}\}\lambda^{r+r'-\gamma(r\wedge r')},$$

where

$$\gamma \quad := \quad \gamma(k, \sigma) := \min\{\tfrac{1}{2}, (\log(\lambda/|\lambda_2|)/\log\lambda\}. \qquad (5.22)$$

Note that, for fixed $k\sigma = \rho$, simple differentiation shows that λ_1 is an increasing function of σ and $|\lambda_2|$ a decreasing function, so that $\lambda_1(\sigma) \geq \lambda_1(0)$, $|\lambda_2(\sigma)| \leq |\lambda_2(0)|$, and hence

$$\frac{\log(\lambda/|\lambda_2|)}{\log\lambda} = 1 - \frac{\log|\lambda_2|}{\log\lambda} \geq 1 - \frac{\log(\sqrt{1+8\rho} - 1)}{\log(\sqrt{1+8\rho} + 1)} \geq \frac{1}{2}$$

in $\rho \leq 1$. Thus, for $\rho \leq 1$, we have $\gamma = \frac{1}{2}$.

Then, from Lemma 5.2 and (5.20) and (5.21), we have the following analogue of Theorem 3.8.

Theorem 5.4 *With the above assumptions and definitions, for $x \in \mathbf{Z}$ and $r = r(x) = \lfloor x/2 \rfloor$, $r' = r'(x) = x - r(x) \leq (x + 1)/2$, we have*

$$\left|\mathbf{P}[D > 2n_d + x] - \mathbf{E} \exp\left\{ -\left(\frac{\lambda}{\lambda - \lambda_2}\right)^2 (\lambda - \sigma)(2\lambda - \sigma)\phi_d^2\lambda^x W_{k,\sigma} W'_{k,\sigma} \right\}\right|$$
$$\leq (\eta'_1(r, r') + \eta'_2(r, r'))(\Lambda\sigma)^{-1/2} + \eta'_3(r, r')(\Lambda\sigma)^{-1/4} + \eta'_4(r, r')(\Lambda\sigma)^{-\gamma/2}.$$

In particular, if $\rho \le 1$,

$$\left|\mathbf{P}[D > 2n_d + x] - \mathbf{E}\exp\left\{-\left(\frac{\lambda}{\lambda-\lambda_2}\right)^2 (\lambda-\sigma)(2\lambda-\sigma)\phi_d^2\lambda^x W_{k,\sigma}W'_{k,\sigma}\right\}\right|$$
$$\le (\eta'_1(r,r') + \eta'_2(r,r'))(\Lambda\sigma)^{-1/2} + (\eta'_3(r,r') + \eta'_4(r,r'))(\Lambda\sigma)^{-1/4}.$$

Note that the expectation in Theorem 5.4 is taken conditional on the initial condition $\hat{M}_0 = \mathbf{e}^{(1)}$.

The theorem can be translated into a uniform bound, similar to that of Theorem 3.9. To do so, we need to be able to control $\mathbf{E}\{e^{-\psi W_{k,\sigma}W'_{k,\sigma}}\}$ for large ψ. The following analogue of Lemma 3.7 makes this possible. To state it, we first need some notation.

For $W_{k,\sigma}$ as in (4.10), let $\phi_{k,\sigma} := (\phi_1, \phi_2)$ denote the Laplace transforms

$$\begin{aligned} \phi_1(\theta) &= \mathbf{E}\{e^{-\theta(f_1^{(1)})^{-1}W_{k,\sigma}} \mid \hat{M}_0 = \mathbf{e}^{(1)}\}; \\ \phi_2(\theta) &= \mathbf{E}\{e^{-\theta(f_1^{(1)})^{-1}W_{k,\sigma}} \mid \hat{M}_0 = \mathbf{e}^{(2)}\} \end{aligned} \tag{5.23}$$

of $\mathcal{L}((f_1^{(1)})^{-1}W_{k,\sigma})$, where $\mathbf{e}^{(i)}$ is the i'th unit vector. Although we now need to distinguish other initial conditions for the branching process, *unconditional* expectations will always in what follows presuppose the initial condition $\hat{M}_0 = \mathbf{e}^{(1)}$, as before. Then, as in Harris [10], p.45, $\phi_{k,\sigma}$ satisfies the Poincaré equation

$$\phi_i(\lambda\theta) \quad = \quad g^i(\phi_1(\theta), \phi_2(\theta)) \quad \text{in} \quad \Re\theta \ge 0; \quad i = 1, 2, \tag{5.24}$$

where g^i is the generating function of $\hat{M}_1$ if $\hat{M}_0 = \mathbf{e}^{(i)}$:

$$g^i(s_1, s_2) = \sum_{r_1, r_2 = 0}^{\infty} p^i(r_1, r_2) s_1^{r_1} s_2^{r_2},$$

where $p^i(r_1, r_2)$ is the probability that an individual of type i has r_1 children of type 1 and r_2 children of type 2. Here, from the binomial structure,

$$g^1(s_1, s_2) \quad = \quad s_2\left(\frac{\sigma}{\Lambda}s_1 + 1 - \frac{\sigma}{\Lambda}\right)^{\Lambda}$$

and

$$g^2(s_1, s_2) \quad = \quad s_2\left(\frac{\sigma}{\Lambda}s_1 + 1 - \frac{\sigma}{\Lambda}\right)^{2k\Lambda}.$$

The Laplace transforms $\phi_{k,\sigma}$ can be bounded as follows.

Lemma 5.5 *For $\theta, \sigma > 0$, we have*

$$\begin{aligned} \phi_{k,\sigma;1}(\theta) =: \phi_1(\theta) &\leq \frac{1}{1+\theta}; \\ \phi_{k,\sigma;2}(\theta) =: \phi_2(\theta) &\leq \frac{1}{1+\theta(\lambda-\sigma)}, \end{aligned}$$

and hence

$$\mathbf{E}\left\{ e^{-\theta\left(f_1^{(1)}\right)^{-2} W_{k,\sigma} W'_{k,\sigma}} \mid \hat{M}(0) = \hat{M}'(0) = \mathbf{e}^{(1)} \right\} \quad \leq \quad \theta^{-1} \log(1+\theta).$$

Proof: We proceed by induction. Put

$$\phi_{i,n}(\theta) \quad = \quad \mathbf{E}\left(e^{-\theta (f_1^{(1)})^{-1} W^{(1)}(n)} \mid \hat{M}(0) = \mathbf{e}^{(i)} \right), \quad i = 1, 2.$$

Then

$$\begin{aligned} \phi_{1,0}(\theta) &= e^{-\theta} \leq \frac{1}{1+\theta}; \\ \phi_{2,0}(\theta) &= e^{-\theta(\lambda-\sigma)} \leq \frac{1}{1+\theta(\lambda-\sigma)}. \end{aligned}$$

Assume that

$$\begin{aligned} \phi_{1,n}(\theta) &\leq \frac{1}{1+\theta}; \\ \phi_{2,n}(\theta) &\leq \frac{1}{1+\theta(\lambda-\sigma)}. \end{aligned}$$

By the Poincaré recursion,

$$\phi_{i,n+1}(\theta) \quad = \quad g^i\left(\phi_{1,n}\left(\frac{\theta}{\lambda}\right), \phi_{2,n}\left(\frac{\theta}{\lambda}\right) \right)$$

for $i = 1, 2$. Hence, using the induction assumption,

$$\begin{aligned} \phi_{1,n+1}(\theta) &\leq \frac{\lambda}{\lambda+\theta(\lambda-\sigma)} \exp\left\{ \sigma\left(\frac{\lambda}{\lambda+\theta} - 1 \right) \right\} \\ &\leq \frac{\lambda}{\lambda(1+\theta)-\theta\sigma} \, \frac{\lambda+\theta}{\lambda+\theta+\theta\sigma} \\ &= \frac{\lambda(\lambda+\theta)}{\lambda(1+\theta)(\lambda+\theta)+\theta^2(\lambda-1-\sigma)\sigma} \\ &\leq \frac{1}{1+\theta}, \end{aligned}$$

and, also from (4.5),

$$\begin{aligned}
\phi_{2,n+1}(\theta) &\le \frac{\lambda}{\lambda+\theta(\lambda-\sigma)} \exp\left\{2k\sigma\left(\frac{\lambda}{\lambda+\theta}-1\right)\right\} \\
&\le \frac{\lambda}{\lambda+\theta+\theta(\lambda-\sigma-1)}\,\frac{\lambda+\theta}{\lambda+\theta+2k\sigma\theta} \\
&= \frac{\lambda}{\lambda+\theta+\theta(\lambda-\sigma-1)}\,\frac{\lambda+\theta}{\lambda+\theta+(\lambda-1)(\lambda-\sigma)\theta} \\
&= \frac{\lambda(\lambda+\theta)}{\lambda(\lambda+\theta)(1+\theta(\lambda-\sigma))+\theta^2(\lambda-1-\sigma)(\lambda-\sigma)(\lambda-1)} \\
&\le \frac{1}{1+\theta(\lambda-\sigma)}.
\end{aligned}$$

Taking limits as $n \to \infty$ proves the first two assertions. The last assertion follows as in Lemma 3.7. []

Theorem 5.6 *Let Δ_d denote a random variable on the integers with distribution given by*

$$\mathbf{P}\left[\Delta_d > x\right] = \mathbf{E}\exp\left\{-\left(\frac{\lambda}{\lambda-\lambda_2}\right)^2 (\lambda-\sigma)(2\lambda-\sigma)\phi_d^2\lambda^x W_{k,\sigma}W'_{k,\sigma}\right\}, \quad x \in \mathbf{Z}, \tag{5.25}$$

and let $D^* = \Delta_d + 2n_d$. Then

$$d_{TV}(\mathcal{L}(D), \mathcal{L}(D^*)) = O\left(\log(\Lambda\sigma)(\Lambda\sigma)^{-\gamma/(4-\gamma)}\right),$$

uniformly in Λ, k and σ such that $k\sigma \le \rho_0$, for any fixed $0 < \rho_0 < \infty$, where γ is given as in (5.22). Hence $d_{TV}(\mathcal{L}(D), \mathcal{L}(D^*)) \to 0$ if $\Lambda\sigma \to \infty$ and $k\sigma$ remains bounded. In particular, $\gamma = 1/2$ if $k\sigma \le 1$, and the approximation error is then of order $O(\log(\Lambda\sigma)(\Lambda\sigma)^{-1/7})$.

PROOF: Fix G, and consider x satisfying $x \le \left\lfloor \frac{G\log(\Lambda\sigma)-2\log\phi_d}{\log\lambda} \right\rfloor$; set $r(x) = \lfloor x/2 \rfloor$, $r'(x) = x - r(x) \le (x+1)/2$. Then it follows from Theorem 5.4 and (5.7) that

$$\begin{aligned}
&|\mathbf{P}[D > 2n_d + x] - \mathbf{P}[D^* > 2n_d + x]| \\
&\le (\eta'_1(r(x), r'(x)) + \eta'_2(r(x), r'(x)))(\Lambda\sigma)^{-1/2}
\end{aligned}$$

$$
\begin{aligned}
&\quad + \eta_3'(r(x), r'(x))(\Lambda\sigma)^{-1/4} + \eta_4'(r(x), r'(x))(\Lambda\sigma)^{-\gamma/2} \\
= \quad & O\Bigg(\lambda^{1/2} \left(\frac{k\sigma \log(\Lambda\sigma)}{\log \lambda} \right)^2 (\Lambda\sigma)^{-(1-3G)/2} \\
&\quad + \lambda^{1/2} \left(\frac{k\sigma \log(\Lambda\sigma)}{\log \lambda} \right) (\Lambda\sigma)^{-(1-G)/2} \\
&\quad + \lambda^{3/4} (\lambda - \sigma)^2 (\Lambda\sigma)^{-(1-3G)/4} \\
&\quad + \lambda^{1-\gamma/2} \left(\frac{\log(\Lambda\sigma)}{\log \lambda} + \big(\sigma^2 \vee (\lambda - 1)\big) \right)^{1/2} (\Lambda\sigma)^{-(\gamma - G(2-\gamma))/2} \Bigg).
\end{aligned}
$$

Also, from Lemma 5.5, recalling that $(f_1^{(1)})^{-2} = \lambda - \sigma$, we have the upper tail estimate

$$
\begin{aligned}
&\mathbf{P}(\Delta_d > x) \\
&\quad \le \quad \frac{(\lambda - \lambda_2)^2}{\lambda^2 (2\lambda - \sigma)} \log \left(1 + \frac{\lambda^2}{(\lambda - \lambda_2)^2} (2\lambda - \sigma)(\Lambda\sigma)^G \right) (\Lambda\sigma)^{-G} \\
&\quad = \quad O\left(\log(\Lambda\sigma)(\Lambda\sigma)^{-G} \right).
\end{aligned}
$$

Comparing the exponents of $\Lambda\sigma$, and remembering that $\gamma \le 1/2$, the best choice of G is $G = \gamma/(4-\gamma)$, making $G = (\gamma - G(2-\gamma))/2$; noting also that $\lambda = O(1 + \sigma + \sqrt{k\sigma})$, the theorem is proved. []

Remembering that the choices $k\sigma = \rho$ and $\Lambda = Lk$ match this model with that of Section 2, we see that $\Lambda\sigma = L\rho$, and that thus Theorem 5.6 matches Theorem 3.9 closely for $\rho \le 1$, but that the total variation distance estimate here becomes bigger as ρ increases. Indeed, if $\rho \to \infty$ and $\sigma = O(k)$, then $\gamma(k, \sigma) \to 0$, and no useful approximation is obtained. This reflects the fact that, when $|\lambda_2|/\lambda$ is close to 1, the martingale $W^{(1)}(n)$ only slowly comes to dominate the behaviour of the two type branching process; for example, from (5.13),

$$
\hat{M}^+(n) = \frac{1}{\lambda - \lambda_2} \left(\frac{\lambda^{n+1}}{f_1^{(1)}} W^{(1)}(n) - \frac{\lambda_2^{n+1}}{f_1^{(2)}} W^{(2)}(n) \right)
$$

then retains a sizeable contribution from $W^{(2)}(n)$ until n becomes extremely large. This is in turn a consequence of taking the shortcuts to have length 1, rather than 0; as a result, the big multiplication, by a factor of 2ρ, occurs only at the *second* time step, inducing substantial fluctuations of period 2 in the branching process, which die away only slowly when ρ is large. However, if $\rho \to \infty$ and $k = O(\sigma^{1-\varepsilon})$ for any $\varepsilon > 0$, then $\liminf \gamma(k,\sigma) > 0$, and it becomes possible for $\mathcal{L}(D)$ and $\mathcal{L}(D^*)$ to be asymptotically close in total variation. This can be deduced from the proof of the theorem by taking $k \sim L^\alpha$ and $\sigma \sim L^{\alpha+\beta}$, for choices of α and β which ensure that σ^2 dominates ρ. Under such circumstances, the effect of two successive multiplications by σ in the branching process dominates that of a single multiplication by 2ρ at the second step, and approximately geometric growth at rate $\lambda \sim \sigma$ results. However, as in all situations in which ρ is a positive power of Λ, interpoint distances are asymptotically bounded, and take one or at most two values with very high probability; an analogue of Corollary 3.10 could for instance also be proved.

If $\rho = k\sigma$ is small, we can again compare the distribution of $W_{k,\sigma}$ with the NE(1) distribution of the limiting variable W in the Yule process (see [7]), using the fact that its Laplace transforms satisfy the Poincaré equation (5.24). Define the operator Ξ by

$$\begin{aligned}
(\Xi\phi)_1(\theta) \quad &:= \quad g^1\left(\phi_1\left(\frac{\theta}{\lambda}\right), \phi_2\left(\frac{\theta}{\lambda}\right)\right) \\
&= \quad \phi_2\left(\frac{\theta}{\lambda}\right)\left(\frac{\sigma}{\Lambda}\phi_1\left(\frac{\theta}{\lambda}\right) + 1 - \frac{\sigma}{\Lambda}\right)^{\Lambda}; \\
(\Xi\phi)_2(\theta) \quad &:= \quad g^2\left(\phi_1\left(\frac{\theta}{\lambda}\right), \phi_2\left(\frac{\theta}{\lambda}\right)\right) \\
&= \quad \phi_2\left(\frac{\theta}{\lambda}\right)\left(\frac{\sigma}{\Lambda}\phi_1\left(\frac{\theta}{\lambda}\right) + 1 - \frac{\sigma}{\Lambda}\right)^{2k\Lambda}.
\end{aligned}$$

Let

$$\mathcal{G} := \left\{\gamma = (\gamma_1, \gamma_2) : [0,\infty)^2 \to \mathbf{R} : \|\gamma\|_{\mathcal{G}} := \sup_{\theta > 0} \max\left\{\frac{|\gamma_1(\theta)|, |\gamma_2(\theta)|}{\theta^2}\right\} < \infty\right\},$$

and

$$\begin{aligned}
\mathcal{H} \quad := \quad &\left\{\psi = (\psi_1, \psi_2) : [0,\infty)^2 \to \mathbf{R} : \frac{\psi_1(\theta) - (1-\theta)}{\theta^2} \text{ is bounded,}\right. \\
&\left. \frac{\psi_2(\theta) - (1 - \theta(\lambda - \sigma))}{\theta^2} \text{ is bounded}\right\}.
\end{aligned}$$

Then $\mathcal{H}$ contains $\phi_{k,\sigma} = (\phi_1, \phi_2)$ as defined in (5.23), since

$$\mathbf{E}\{(f_1^{(1)})^{-1} W_{k,\sigma} \mid \hat{M}(0) = \mathbf{e}^{(1)}\} = 1; \quad \mathbf{E}\{(f_1^{(1)})^{-1} W_{k,\sigma} \mid \hat{M}(0) = \mathbf{e}^{(2)}\} = \lambda - \sigma,$$

and taking limits in (5.3) shows that $\operatorname{Var} W_{k,\sigma}$ exists. We next show that Ξ is a contraction on $\mathcal{H}$.

Lemma 5.7 *The operator Ξ is a contraction on $\mathcal{H}$, and, for all $\psi, \chi \in \mathcal{H}$,*

$$\|\Xi\psi - \Xi\chi\|_{\mathcal{G}} \leq \left(\frac{2k\sigma+1}{\lambda^2}\right)\|\psi-\chi\|_{\mathcal{G}}.$$

Remark. Note that

$$\frac{2k\sigma+1}{\lambda^2} = \frac{\lambda^2-(\lambda-1)(\sigma+1)}{\lambda^2} < 1.$$

Proof: For all $\psi, \chi \in \mathcal{H}$ and $\theta > 0$, observe that $\psi - \chi \in \mathcal{G}$. We then compute

$$\begin{aligned} |(\Xi\psi)_1(\theta) - (\Xi\chi)_1(\theta)| &\leq \left|\psi_2\left(\frac{\theta}{\lambda}\right) - \chi_2\left(\frac{\theta}{\lambda}\right)\right| \\ &\quad + \sigma\left|\psi_1\left(\frac{\theta}{\lambda}\right) - \chi_1\left(\frac{\theta}{\lambda}\right)\right|, \end{aligned}$$

so that

$$\begin{aligned} \frac{|(\Xi\psi)_1(\theta) - (\Xi\chi)_1(\theta)|}{\theta^2} &\leq \frac{1}{\lambda^2}\frac{\left|\psi_2\left(\frac{\theta}{\lambda}\right) - \chi_2\left(\frac{\theta}{\lambda}\right)\right|}{\left(\frac{\theta}{\lambda}\right)^2} \\ &\quad + \frac{\sigma}{\lambda^2}\frac{\left|\psi_1\left(\frac{\theta}{\lambda}\right) - \chi_1\left(\frac{\theta}{\lambda}\right)\right|}{\left(\frac{\theta}{\lambda}\right)^2} \\ &\leq \frac{\sigma+1}{\lambda^2}\|\psi-\chi\|_{\mathcal{G}}. \end{aligned} \tag{5.26}$$

Similarly,

$$\begin{aligned} |(\Xi\psi)_2(\theta) - (\Xi\chi)_2(\theta)| &\leq \left|\psi_2\left(\frac{\theta}{\lambda}\right) - \chi_2\left(\frac{\theta}{\lambda}\right)\right| \\ &\quad + 2k\sigma\left|\psi_1\left(\frac{\theta}{\lambda}\right) - \chi_1\left(\frac{\theta}{\lambda}\right)\right|, \end{aligned}$$

and

$$\frac{|(\Xi\psi)_2(\theta) - (\Xi\chi)_2(\theta)|}{\theta^2} \leq \left(\frac{2k\sigma+1}{\lambda^2}\right)\|\psi-\chi\|_{\mathcal{G}}.$$

Taking the maximum of the bounds finishes the proof. []

Thus, for any starting function $\psi = (\psi_1, \psi_2) \in \mathcal{H}$ and for $\phi_{k,\sigma} = (\phi_1, \phi_2)$ given in (5.23), we have

$$\begin{aligned} \|\phi_{k,\sigma} - \psi\|_{\mathcal{G}} &\leq \|\Xi\phi_{k,\sigma} - \Xi\psi\|_{\mathcal{G}} + \|\Xi\psi - \psi\|_{\mathcal{G}} \\ &\leq \frac{2k\sigma+1}{\lambda^2}\|\phi_{k,\sigma} - \psi\|_{\mathcal{G}} + \|\Xi\psi - \psi\|_{\mathcal{G}}, \end{aligned}$$

so that

$$\|\phi_{k,\sigma} - \psi\|_{\mathcal{G}} \quad \leq \quad \frac{\lambda^2}{\lambda^2 - (2k\sigma + 1)} \|\Xi\psi - \psi\|_{\mathcal{G}}. \tag{5.27}$$

Hence a function ψ such that $\|\Xi\psi - \psi\|_{\mathcal{G}}$ is small provides a good approximation to $\phi_{k,\sigma}$.

As a candidate ψ, we try

$$\begin{aligned} \psi_{(1)}(\theta) &= \frac{1}{1+\theta}, \\ \psi_{(2)}(\theta) &= \frac{1}{1+\theta(\lambda-\sigma)}; \end{aligned} \tag{5.28}$$

Lemma 5.5 shows that this pair dominates $\phi_{k,\sigma}$.

Lemma 5.8 *For ψ given in (5.28), we have*

$$\|\Xi\psi - \psi\|_{\mathcal{G}} \leq \frac{2k\sigma(\lambda^2 - \lambda\sigma - 1 + k\sigma)}{\lambda^2}.$$

PROOF: For $\theta > 0$, we have

$$\begin{aligned} &(\Xi\psi)_1(\theta) - \psi_1(\theta) \\ &\quad = \frac{\lambda}{\lambda + \theta(\lambda - \sigma)} \left(1 - \frac{\sigma\theta}{\Lambda(\lambda + \theta)}\right)^{\Lambda} - \frac{1}{1+\theta} \\ &\quad = \frac{1}{1+\theta} \left\{ \frac{\lambda(1+\theta)}{\lambda + \theta(\lambda - \sigma)} \left(1 - \frac{\sigma\theta}{(\lambda+\theta)}\right) - 1 \right\} + R_1, \end{aligned}$$

where

$$R_1 \quad = \quad \frac{\lambda}{\lambda + \theta(\lambda-\sigma)} \left[\left(1 - \frac{\sigma\theta}{\Lambda(\lambda+\theta)}\right)^{\Lambda} - 1 + \frac{\sigma\theta}{\lambda+\theta} \right].$$

Moreover,

$$\begin{aligned} &\frac{1}{1+\theta} \left\{ \frac{\lambda(1+\theta)}{\lambda + \theta(\lambda-\sigma)} \left(1 - \frac{\sigma\theta}{\lambda+\theta}\right) - 1 \right\} \\ &\qquad = \frac{\theta^2\sigma(1-\lambda)}{(1+\theta)(\lambda + (\lambda-\sigma)\theta)(\lambda+\theta)}. \end{aligned}$$

From Taylor's expansion, it follows that

$$\begin{aligned} |R_1| &\leq \frac{\lambda\Lambda(\Lambda-1)\sigma^2\theta^2}{2(\lambda + (\lambda-\sigma)\theta)\,\Lambda^2(\lambda+\theta)^2} \\ &\leq \frac{\sigma^2\theta^2}{2\lambda^2}. \end{aligned}$$

Hence

$$\frac{|(\Xi\psi)_1(\theta) - \psi_1(\theta)|}{\theta^2} \quad \leq \quad \frac{\sigma(2(\lambda-1)+\sigma)}{2\lambda^2}. \tag{5.29}$$

Similarly,

$$\begin{aligned} &(\Xi\psi)_2(\theta) - \psi_2(\theta) \\ &\quad = \frac{\lambda}{\lambda+\theta(\lambda-\sigma)}\left(1-\frac{\sigma\theta}{\Lambda(\lambda+\theta)}\right)^{2k\Lambda} - \frac{1}{1+\theta(\lambda-\sigma)} \\ &\quad = \frac{1}{1+\theta(\lambda-\sigma)}\left\{\frac{\lambda(1+\theta(\lambda-\sigma))}{\lambda+\theta(\lambda-\sigma)}\left(1-\frac{2k\sigma\theta}{\lambda+\theta}\right)-1\right\} + R_2, \end{aligned}$$

where

$$R_2 \quad = \quad \frac{\lambda}{\lambda+\theta(\lambda-\sigma)}\left[\left(1-\frac{\sigma\theta}{\Lambda(\lambda+\theta)}\right)^{2k\Lambda} - 1 + \frac{2k\sigma\theta}{\lambda+\theta}\right].$$

Using (4.5), we obtain

$$\begin{aligned} &\frac{1}{1+\theta(\lambda-\sigma)}\left\{\frac{\lambda(1+\theta(\lambda-\sigma))}{\lambda+\theta(\lambda-\sigma)}\left(1-\frac{2k\sigma\theta}{\lambda+\theta}\right)-1\right\} \\ &\quad = \frac{\theta^2(\lambda-1)(\lambda-\sigma)(1+\lambda\sigma-\lambda^2)}{(1+\theta(\lambda-\sigma))(\lambda+(\lambda-\sigma)\theta)(\lambda+\theta)} \\ &\quad = \frac{2k\sigma\theta^2(1+\lambda\sigma-\lambda^2)}{(1+\theta(\lambda-\sigma))(\lambda+(\lambda-\sigma)\theta)(\lambda+\theta)}. \end{aligned}$$

From Taylor's expansion, it now follows that

$$\begin{aligned} |R_2| &\leq \frac{2k\Lambda(2k\Lambda-1)\lambda\sigma^2\theta^2}{2(\lambda+(\lambda-\sigma)\theta)\,\Lambda^2(\lambda+\theta)^2} \\ &\leq \frac{2k^2\sigma^2\theta^2}{\lambda^2}. \end{aligned}$$

Hence

$$\frac{|(\Xi\psi)_2(\theta) - \psi_2(\theta)|}{\theta^2} \quad \leq \quad \frac{2k\sigma(\lambda^2-\lambda\sigma-1+k\sigma)}{\lambda^2},$$

completing the proof, since

$$2(\lambda^2-\lambda\sigma-1+k\sigma) = 2(\lambda-1)+2(3k-1)\sigma > 2(\lambda-1)+\sigma.$$

[]

This enables us to prove the exponential approximation to $\mathcal{L}(W_{k,\sigma})$ when $k\sigma$ is small.

Theorem 5.9 *As $k\sigma \to 0$, $\mathcal{L}(W_{k,\sigma}) \to \mathrm{NE}(1)$.*

PROOF: Let $\phi_{k,\sigma}$ be as in (5.23), and ψ as in (5.28). Then $(\phi_{k,\sigma})_1$ is the Laplace transform of

$$\mathcal{L}((f_1^{(1)})^{-1}W_{k,\sigma}) := \mathcal{L}((f_1^{(1)})^{-1}W_{k,\sigma} \mid \hat{M}(0) = \mathbf{e}^{(1)}),$$

and ψ_1 that of NE(1), and $(f_1^{(1)})^{-1} = (\lambda - \sigma)^{1/2} \to 1$ as $k\sigma \to 0$. Hence it is enough to show that

$$\lim_{k\sigma \to 0} \|\phi_{k,\sigma} - \psi\|_{\mathcal{G}} = 0.$$

However, using Lemma 5.8 and (5.27), we obtain

$$\begin{aligned} &\|\phi_{k,\sigma} - \psi\|_{\mathcal{G}} \\ &\quad \le \left(\frac{\lambda^2}{\lambda^2 - 1 - 2k\sigma}\right) \|\Xi\psi - \psi\|_{\mathcal{G}} \\ &\quad \le 2k\sigma \, \frac{\lambda^2 - \lambda\sigma - 1 + k\sigma}{\lambda^2 - 1 - 2k\sigma} \\ &\quad \le 2k\sigma \, \frac{k\sigma(5 + 2\sigma)}{2k\sigma(\sigma + 1)} \, (1 + 2k\sigma(\sigma + 1)) \\ &\quad \le k\sigma(1 + 2k\sigma)(5 + 2\sigma) \to 0, \end{aligned} \tag{5.30}$$

since $k\sigma \to 0$. This proves the theorem. []

Again we can use this result to derive an approximation to the distribution of the distance for D, based on a corresponding distribution derived from the NE(1) distribution. The starting point is the following result.

Theorem 5.10 *Let W, W' be independent NE(1) variables. Then, for all $\theta > 0$,*

$$\begin{aligned} &\left| \mathbf{E} \exp\left\{-\theta(\lambda - \sigma)W_{k,\sigma}W'_{k,\sigma}\right\} - \mathbf{E}e^{-\theta W W'} \right| \\ &\quad \le \quad 5\theta^2 k\sigma \left\{(3\sigma/2) + (1 + 2k\sigma)(5 + 2\sigma)\right\}. \end{aligned}$$

PROOF: As in the proof of Theorem 3.14, with $\phi_{k,\sigma}$ as in (5.23) and ψ as in (5.28), and because, from (4.8), $\lambda - \sigma = (f_1^{(1)})^{-2}$, we have

$$\mathbf{E} \exp\left\{-\theta(\lambda - \sigma)W_{k,\sigma}W'_{k,\sigma}\right\} - \mathbf{E}e^{-\theta W W'}$$

$$
\begin{aligned}
&= \mathbf{E}\left\{(\phi_{k,\sigma})_1(\theta(f_1^{(1)})^{-1}W'_{k,\sigma})\right\} - \mathbf{E}\psi_1(\theta W) \\
&= \mathbf{E}\left\{(\Xi\phi_{k,\sigma})_1(\theta(f_1^{(1)})^{-1}W'_{k,\sigma})\right\} - \mathbf{E}\left\{(\Xi\psi)_1(\theta(f_1^{(1)})^{-1}W'_{k,\sigma})\right\} \\
&\quad +\mathbf{E}\left\{(\Xi\psi)_1(\theta(f_1^{(1)})^{-1}W'_{k,\sigma})\right\} - \mathbf{E}\left\{\psi_1(\theta(f_1^{(1)})^{-1}W'_{k,\sigma})\right\} \\
&\quad +\mathbf{E}\left\{\psi_1(\theta(f_1^{(1)})^{-1}W'_{k,\sigma})\right\} - \mathbf{E}\psi_1(\theta W).
\end{aligned}
$$

Now (5.26) in the proof of Lemma 5.7 gives

$$
\begin{aligned}
&\left|\mathbf{E}\left\{(\Xi\phi)_1(\theta(f_1^{(1)})^{-1}W'_{k,\sigma})\right\} - \mathbf{E}\left\{(\Xi\psi_{k,\sigma})_1(\theta(f_1^{(1)})^{-1}W'_{k,\sigma})\right\}\right| \\
&\quad \le \theta^2(f_1^{(1)})^{-2}\frac{\sigma+1}{\lambda^2}\,\|\phi_{k,\sigma}-\psi\|_{\mathcal{G}}\mathbf{E}\{(W'_{k,\sigma})^2\} \\
&\quad \le \theta^2(1+\kappa^2)\frac{\sigma+1}{\lambda^2}\,\|\phi_{k,\sigma}-\psi\|_{\mathcal{G}} \\
&\quad \le 3\theta^2\frac{\sigma+1}{\lambda^2}\,\|\phi_{k,\sigma}-\psi\|_{\mathcal{G}},
\end{aligned}
$$

from (4.10), (5.3) and (5.7), and (5.30) then implies that the above expression can be bounded by

$$
3\theta^2\,\frac{\sigma+1}{\lambda^2}\,k\sigma(1+2k\sigma)(5+2\sigma).
$$

Similarly, from (5.29) in the proof of Lemma 5.8,

$$
\left|\mathbf{E}\left\{(\Xi\psi)_1(\theta(f_1^{(1)})^{-1}W'_{k,\sigma})\right\} - \mathbf{E}\left\{\psi_1(\theta(f_1^{(1)})^{-1}W'_{k,\sigma})\right\}\right| \le 3\theta^2\,\frac{\sigma\{2(\lambda-1)+\sigma\}}{2\lambda^2}.
$$

Then, with $W \sim \mathrm{NE}\,(1)$ independent of $W'_{k,\sigma}$, we have

$$
\mathbf{E}\left\{\psi_1(\theta(f_1^{(1)})^{-1}W'_{k,\sigma})\right\} = \mathbf{E}\left\{e^{-\theta(f_1^{(1)})^{-1}WW'_{k,\sigma}}\right\} = \mathbf{E}\left\{(\phi_{k,\sigma})_1(\theta W)\right\},
$$

and hence, from (5.30) in the proof of Theorem 5.9, it follows that

$$
\begin{aligned}
\left|\mathbf{E}\left\{\psi_1(\theta(f_1^{(1)})^{-1}W'_{k,\sigma})\right\} - \mathbf{E}\psi_1(\theta W)\right| &= \left|\mathbf{E}\left\{(\phi_{k,\sigma})_1(\theta W)\right\} - \mathbf{E}\psi_1(\theta W)\right| \\
&\le 2\theta^2 k\sigma(1+2k\sigma)(5+2\sigma).
\end{aligned}
$$

Since $\lambda - 1 < 2k\sigma$ and $\lambda^2 > \sigma + 1$, the assertion follows from the triangle inequality. []

Recalling that

$$
\mathbf{E}e^{-\theta WW'} = \int_0^\infty \frac{e^{-y}}{1+\theta y}\,dy,
$$

we can now derive the analogue of Theorem 3.15.

Theorem 5.11 *As in Theorem 5.6, let Δ_d denote a random variable on the integers with distribution given by*

$$\mathbf{P}\left[\Delta_d > x\right] = \mathbf{E}\exp\left\{-\left(\frac{\lambda}{\lambda-\lambda_2}\right)^2(\lambda-\sigma)(2\lambda-\sigma)\phi_d^2\lambda^x W_{k,\sigma}W'_{k,\sigma}\right\}, \quad x \in \mathbf{Z}.$$

Let T denote a random variable on $\mathbf{R}$ with distribution given by

$$\mathbf{P}\left[T > z\right] = \int_0^\infty \frac{e^{-y}}{1+2ye^{2z}}\,dy.$$

Then

$$\sup_{z\in\mathbf{R}}\left|\mathbf{P}\left[\tfrac{\lambda-1}{2}\Delta_d > z\right] - \mathbf{P}[T > z]\right| \quad = \quad O\left\{(k\sigma)^{1/3}\left(1+\log\left(1/k\sigma\right)\right)\right\},$$

uniformly in $k\sigma \leq 2$.

Proof: We use an argument similar to the proof of Theorem 3.15. Putting

$$\tilde{c}(\lambda) = \frac{\log\lambda}{\lambda-1} = \frac{\log(1+2\,\frac{\lambda-1}{2})}{2\,\frac{\lambda-1}{2}},$$

we have, as before,

$$1 \geq \tilde{c}(\lambda) \geq 1 - \frac{\lambda-1}{2};$$

we also write

$$\beta(\lambda) \quad := \quad \left(\frac{\lambda}{\lambda-\lambda_2}\right)^2\left(\lambda-\frac{\sigma}{2}\right)\phi_d^2.$$

We use the following bounds. First, from the characteristic equation (4.5), we have

$$0 \;\leq\; \lambda - 1 \quad = \quad \frac{2k\sigma}{\lambda-\sigma} \;\leq\; 2k\sigma.$$

Then, since $\sqrt{a+b} \leq \sqrt{a}+\sqrt{b}$ for $a, b \geq 0$, it follows that

$$1+\sigma \;\leq\; \lambda \quad\leq\quad 1+\sigma+\sqrt{\sigma(2k-1)}.$$

Thus, and because $\lambda^{-1} \leq \phi_d \leq 1$ and $\lambda_2 < 0$, we have

$$\beta(\lambda) \quad\leq\quad 1+(\sigma/2)+\sqrt{\sigma(2k-1)}$$

and

$$\begin{aligned}\beta(\lambda) &\geq \left(\frac{1}{\lambda-\lambda_2}\right)^2 \\ &= 1-\frac{2\sigma(4k-1)+\sigma^2}{(\sigma+1)^2+4\sigma(2k-1)}.\end{aligned}$$

This in turn gives

$$\begin{aligned}|\beta(\lambda)-1| &\leq \max\left\{(\sigma/2)+\sqrt{\sigma(2k-1)}, \frac{2\sigma(4k-1)+\sigma^2}{(\sigma+1)^2+4\sigma(2k-1)}\right\} \\ &=: \Gamma(\sigma,k) \\ &= O\left(\max\{\sigma,\sqrt{k\sigma}\}\right).\end{aligned}$$

For the main part of the distribution, we write

$$\mathbf{P}\left[\tfrac{\lambda-1}{2}\Delta_d > z\right] - \mathbf{P}[T>z]$$
$$= \mathbf{P}\left[\tfrac{\lambda-1}{2}\Delta_d > z\right] - \int_0^\infty e^{-y}\left(1+2y\beta(\lambda)e^{2z\tilde{c}(\lambda)}\right)^{-1} dy \tag{5.31}$$
$$+ \int_0^\infty e^{-y}\left(1+2y\beta(\lambda)e^{2z\tilde{c}(\lambda)}\right)^{-1} dy - \mathbf{P}[T>z\tilde{c}(\lambda)] \tag{5.32}$$
$$+ \mathbf{P}[T>z\tilde{c}(\lambda)] - \mathbf{P}[T>z]. \tag{5.33}$$

Now, for (5.31), Theorem 5.10 yields

$$\begin{aligned}&\left|\mathbf{P}\left[\tfrac{\lambda-1}{2}\Delta_d > z\right] - \int_0^\infty e^{-y}\left(1+2y\beta(\lambda)e^{2z\tilde{c}(\lambda)}\right)^{-1} dy\right| \\ &= \left|\mathbf{E}\exp\left\{-2\beta(\lambda)e^{2z\tilde{c}(\lambda)}W_{k,\sigma}W'_{k,\sigma}\right\} - \int_0^\infty e^{-y}\left(1+2y\beta(\lambda)e^{2z\tilde{c}(\lambda)}\right)^{-1} dy\right| \\ &\leq 4\beta(\lambda)^2 e^{4z\tilde{c}(\lambda)} 5k\sigma\left\{(3\sigma/2)+(1+2k\sigma)(5+2\sigma)\right\} \\ &\leq (2\lambda-\sigma)^2 e^{4z} 5k\sigma\left\{(3\sigma/2)+(1+2k\sigma)(5+2\sigma)\right\}.\end{aligned}$$

With (3.18), we have, for (5.32), that

$$\begin{aligned}&\left|\int_0^\infty e^{-y}\left(1+2y\beta(\lambda)e^{2z\tilde{c}(\lambda)}\right)^{-1} dy - \mathbf{P}[T>z\tilde{c}(\lambda)]\right| \\ &\leq 2e^{2z\tilde{c}(\lambda)}\frac{|\beta(\lambda)-1|}{\max\{1, 2\beta(\lambda)e^{2z\tilde{c}(\lambda)}, 2e^{2z\tilde{c}(\lambda)}\}} \\ &\leq 2\frac{e^{2z\tilde{c}(\lambda)}}{\max\{1, 2e^{2z\tilde{c}(\lambda)}\}}\,|\beta(\lambda)-1| \\ &\leq \Gamma(\sigma,k).\end{aligned}$$

Similarly, for (5.33), because $1-\tilde{c}(\lambda) \le \frac{\lambda-1}{2} \le k\sigma$ and from Taylor's expansion, it follows that

$$\begin{aligned}
&|\mathbf{P}[T > z\tilde{c}(\lambda)] - \mathbf{P}[T > z]| \\
&= \left| \int_0^\infty e^{-y} \left(1 + 2ye^{2z\tilde{c}(\lambda)}\right)^{-1} dy - \int_0^\infty e^{-y} \left(1 + 2ye^{2z}\right)^{-1} dy \right| \\
&\le 2e^{2z} \frac{|e^{-2z(1-\tilde{c}(\lambda))} - 1|}{\max\{1, 2e^{2z\tilde{c}(\lambda)}, 2e^{2z}\}}.
\end{aligned}$$

If $z > 0$, this gives

$$|\mathbf{P}[T > z\tilde{c}(\lambda)] - \mathbf{P}[T > z]| \le 2z(1-\tilde{c}(\lambda)) \le 2k\sigma z;$$

if $z \le 0$, we have

$$|\mathbf{P}[T > z\tilde{c}(\lambda)] - \mathbf{P}[T > z]| \quad \le \quad 2|z|(1-\tilde{c}(\lambda))e^{-2z\tilde{c}(\lambda)} \ \le \ 2k\sigma|z|e^{2z(1-k\sigma)}.$$

Hence we conclude that, uniformly in $k\sigma \le 1/2$,

$$\begin{aligned}
&\mathbf{P}\left[\tfrac{\lambda-1}{2}\Delta_d > z\right] - \mathbf{P}[T > z] \\
&\le 5k\sigma e^{4z}(2\lambda-\sigma)^2 \{(3\sigma/2) + (1+2k\sigma)(5+2\sigma)\} \\
&\qquad + \Gamma(\sigma, k) + 2k\sigma|z| \min\{1, e^{2z(1-k\sigma)}\} \\
&\le C_1 \left\{ k\sigma(e^{4z}+1) + \sqrt{k\sigma} \right\}, \qquad (5.34)
\end{aligned}$$

for some constant C_1.

For the large values of z, where the bound given in (5.34) becomes useless, we can estimate the upper tails of the random variables separately. First, for $x \in \mathbf{Z}$, we have

$$\mathbf{P}[\Delta_d > x] = \mathbf{E} \exp\left\{ -2\beta(\lambda)\lambda^x (f_1^{(1)})^{-2} W_{k,\sigma} W'_{k,\sigma} \right\},$$

so that, by Lemma 5.5, it follows that

$$\begin{aligned}
&\mathbf{P}\left[\tfrac{\lambda-1}{2}\Delta_d > z\right] \\
&= \mathbf{E}\exp\left\{ -2\beta(\lambda)e^{2z\tilde{c}(\lambda)} (f_1^{(1)})^{-2} W_{k,\sigma} W'_{k,\sigma} \right\} \\
&\le \left(\frac{\lambda-\lambda_2}{\lambda}\right)^2 (2\lambda-\sigma)^{-1} \phi_d^{-2} e^{-2z\tilde{c}(\lambda)} \log\left(1 + \left(\frac{\lambda}{\lambda-\lambda_2}\right)^2 (2\lambda-\sigma)\phi_d^2 e^{2z\tilde{c}(\lambda)}\right) \\
&\le 4\lambda e^{-2z\tilde{c}(\lambda)} \log\left(1 + (2\lambda-\sigma)e^{2z\tilde{c}(\lambda)}\right) \\
&\le 4\lambda e^{-2z(1-k\sigma)} \log\left(1 + \left(2+\sigma+2\sqrt{\sigma(2k-1)}\right) e^{2z}\right), \qquad z \in \tfrac{\lambda-1}{2}\mathbf{Z}.
\end{aligned}$$

For the upper tail of T, as in (3.16) and with $z > 0$, we have

$$\mathbf{P}[T > z\tilde{c}(\lambda)] \quad \leq \quad e^{-2z\tilde{c}(\lambda)}(1 + z\tilde{c}(\lambda)) \ \leq \ (1 + z)e^{-2z(1-k\sigma)}.$$

Combining these two tail estimates, we find that, for $z > 0$,

$$|\mathbf{P}[T > z\tilde{c}(\lambda)] - \mathbf{P}[T > z]| \quad \leq \quad C_2(1 + z)e^{-2z(1-\sigma k)}, \qquad (5.35)$$

uniformly in $k\sigma \leq 1/2$, for some constant C_2. Applying the bound (5.34) when $z \leq (6 - 2k\sigma)^{-1}\log(1/k\sigma)$ and (5.35) for all larger z, and remembering that T has bounded density, so that the discrete nature of Δ_d requires only a small enough correction, a bound of the required order follows. □

Acknowledgement. We would like to thank Svante Janson for very helpful discussions concerning the comparison with [13].

References

[1] Albert, R. and Barabasi, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics* **74**, 47–97.

[2] Asmussen, S. and Hering, H. (1983). *Branching Processes.* Birkhäuser, Boston, Basel, Stuttgart.

[3] Athreya, K.B. and Ney, P.E. (1972). *Branching Processes.* Springer, Berlin, Heidelberg, New York.

[4] Ball, F., Mollison, D. and Scalia-Tomba, G. (1997). Epidemics with two levels of mixing. *Ann. Appl. Probab.* **7**, 46–89.

[5] Barabasi, A.-L. (2002). *Linked: the new science of networks.* Perseus, Cambridge, Massachusetts.

[6] Barbour, A.D., Holst, L., and Janson, S. (1992). *Poisson Approximation.* Oxford Science Publications.

[7] Barbour, A.D. and Reinert, G. (2001). Small Worlds. *Random Structures and Algorithms* **19**, 54–74.

[8] Dorogovtsev, S.N. and Mendes, J.F.F. (2003). *Evolution of Networks: From Biological Nets to the Internet and WWW.* Oxford University Press, Oxford.

[9] Gumbel, E.J. (1958). *Statistics of extremes.* Columbia University Press.

[10] Harris, T.E. (1989). *The Theory of Branching Processes.* Dover, New York.

[11] Moore, C. and Newman, M.E.J. (2000). Epidemics and percolation in small-world networks. *Phys. Rev. E.* **61**, 5678–5682.

[12] Newman, M.E.J. and Watts, D.J. (1999). Scaling and percolation in the small-world network model. *Phys. Rev. E*, **60**, 7332-7344.

[13] Newman, M.E.J., Moore, C. and Watts, D.J. (2000). Mean-field solution of the small-world network model. *Phys. Rev. Lett.* **84**, 3201–3204.

[14] Watts, D.J. (1999). *Small Worlds.* Princeton University Press.

[15] Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of "small-world" networks. *Nature* **393**, 440–442.