Discrete small world networks

A. D. Barbour* and G. Reinert†
Universität Zürich and University of Oxford

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.

Keywords: Small-world networks, shortest path length, branching process
AMS subject classification: 90B15, 60J85

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 $\frac{1}{2}L\rho$ of shortcuts tends to infinity;

*Angewandte Mathematik, Universität Zürich, Winterthurertrasse 190, CH-8057 ZÜRICH; ADB was supported in part by Schweizerischer Nationalfonds Projekt Nr. 20-67909.02.

†Department of Statistics, University of Oxford, 1 South Parks Road, OXFORD OX1 3TG, UK; GR was supported in part by EPSRC grant no. GR/R52183/01.

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 d_K 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.

The main idea is to find the shortest distance between two randomly chosen points P and P' by considering the sets of points $R(n)$ and $R'(n)$ that can be reached within distance n from P and P', respectively, as n increases. The value of n at which $R(n)$ and $R'(n)$ first intersect then gives half the shortest path between P and P'.

Each of the sets $R(n)$ and $R'(n)$ consists of a union of intervals, which grow in size as n increases. Their numbers may increase, because of new intervals added through shortcuts, and they may decrease, if pairs of intervals grow into one another. This makes their evolution rather complicated to analyze. However, in the tradition of branching process approximations in epidemic theory, there is a simpler process with branching structure which acts as a good enough approximation until the value of n at which the first intersection takes place, and with which we can work. Conditional on the number and sizes of the intervals in this process, the number of intersections has a distribution which is close to Poisson, and removing the conditioning leads to a mixed Poisson approximation for the number of intersections. To calculate the mixing distribution, and thence to characterize the limiting distribution of the shortest path length, the existence of the martingale limit for the branching process and its limiting distribution play a crucial role.

In Section 2, we introduce the continuous model in discrete time. Here, the branching approximation is not a Yule process, as was the case for the continuous circle [NMW] model treated in [7], and the distribution of its martingale limit is not available in closed form; as a result, our limiting approximation to the shortest path length is not as explicit as it was there. Another main difference from the [NMW] model is that the limiting distribution of the shortest path length may be concentrated on just one or two points, when the probability of shortcuts is large; the distribution is discrete but more widely spread if the probability of shortcuts is moderate; and only if the probability of shortcuts is very small is the distribution close to the continuous limiting distribution of the [NMW] model, for which [7] gives a closed-form expression.

The proofs of these results are to be found in Section 3. In Section 3.1, the mixed Poisson approximation for the number of intersections in the approximating branching process is given in Proposition 3.1. This is converted in Corollary 3.4 to an approximation to the probability that the two original processes started at P and P' have not yet met. The approximation is good for the times that we are interested in. In Section 3.2, we give a limiting distribution for the shortest path length in Corollary 3.6, and a uniform bound for the distributional approximation in Theorem 3.10. Finally, in Section 3.4, we show that the limiting distribution in the [NMW] model is recovered, if the probability of shortcuts is very small. For this, we employ a contraction argument to show that the Laplace transforms of the two limiting distributions are close. The resulting bound on the Kolmogorov distance between the limiting distribution in our model and that for the [NMW] model is given in Theorem 3.16.

Section 4 describes the discrete circle model in discrete time, which is the usual small-world model, and gives the main results. Shortcuts now have length 1 instead of length 0, and so a newly created interval, which consists of a single point, makes relatively little contribution to the creation of new intervals at the next time step; there are only the possible shortcuts from a single point available for this, whereas established intervals typically have $2k$ points to act as sources of potential shortcuts. This leads to a hesitation in the growth of the process, with significant consequences in the large ρ regime. Mathematically, the analysis becomes more complicated, because we now have to consider a two-type branching process, the types corresponding to newly created and established intervals. The largest eigenvalue and corresponding eigenvector of the branching process mean matrix now play a major role in the martingale limit and in the mixed Poisson approximation, and hence in the limiting distribution of the shortest path. Again, this limiting distribution may be concentrated on just one or two points if the probability of shortcuts is large; the distribution is discrete but spread out if the probability of shortcuts is intermediate; and it approaches the limiting distribution of the [NMW] model only if the probability of shortcuts is very small.

Section 5 gives the proofs; its structure is similar to that of Section 3. Lemma 5.4 in Section 5.1 gives the mixed Poisson approximation for the probability that there are no intersections, and Theorem 5.9 in Subsection 5.2 converts the result into a uniform approximation for the distribution of the shortest path length. In Subsection 5.3, to assess the distance to the limiting distribution for the [NMW] model, we again employ Laplace transforms and a contraction argument, leading to the uniform bound in Theorem 5.14.

It should be noted that, in all our results, the probability of shortcuts and the neighbourhood size are both allowed to vary with the circumference of the circle. Our approximations come complete with error bounds, which are explicitly expressed in terms of these quantities. Limiting behaviour is investigated under conditions in which the expected number of shortcuts grows to infinity.

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 $\mathrm{Po}\,(L\rho/2)$ distributed number of uniform and independent random chords, but now with a new measure of distance $d(P, P')$ between points P and P'. This distance is the minimum of $d(\gamma)$ over paths γ along the graph between P and P', 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 interval 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 results in a new union $R'(n+1)$ of $B'(n+1)$ intervals; note that $B'(n+1)$ may be less than $B(n)$.

Then, to get from $R'(n+1)$ to $R(n+1)$, we need to add any new one point intervals which are connected to $\partial R(n+1) := R'(n+1) \setminus R(n)$ by shortcuts (since these are taken to have length zero). These arise principally from those chords whose label 1 end points lie in $\partial R(n+1)$, and whose label 2 end points lie outside $R'(n+1)$; thus shortcuts are in essence those chords whose label 1 end points are reached before their label 2 end points. However, if both ends of a chord are first reached at the same time point, then we flip a fair coin to decide whether or not to accept it. Thus, for each $q \in \partial R(n+1) \cap Q$, we accept the chord $\{q, q'\}$ if $q' = q'(q) \notin R'(n+1)$, 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 $Q(n+1) := \{q' \notin R'(n+1) : \{q, q'\} \text{ newly accepted}\}$, 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.

We shall actually use the construction with different initial conditions. We shall start with $R(0) = \{P, P'\}$ and $B(0) = 2$, to realize the sets of points accessible from either of the two points $P, P' \in C$ within distance n, for each $n \geq 1$; and with $R(0) = P$ and $B(0) = 1$, but then adding the point P' to $R(1)$ and increasing $B(1)$ by 1, in order to realize the sets of points accessible either from P within distance n or from P' within distance $n-1$, for each $n \geq 1$. In either case, we can also record, at each time n, the information as to which intervals are accessible from P and which from P' within the allotted distances. If, at time n, there is an interval which is accessible from both P and P', then there are some points in it (not necessarily all of them) which can be reached from both P and P'; this implies that $d(P, P') \leq 2n$ with the first initial condition, and that $d(P, P') \leq 2n-1$ with the second. If not, then $d(P, P') > 2n$ or $d(P, P') > 2n-1$, respectively. Thus whether or not $d(P, P') \leq k$ is true for any particular k, even or odd, can be determined by knowing whether, at the appropriate time step and with the appropriate initial condition, there are intervals accessible from both P and P'.

For our analysis, as in [7], we define a closely related birth and growth process $S^*(n)$, starting from the same initial conditions, using the same set of potential chords, and having the same unit growth per time step. The differences are that *every* potential chord is included in S^*, so that no thinning takes place, and the chords that were not accepted for R initiate independent birth and growth processes having the same distribution as S^*, starting from their label 2 end points. Additionally, whenever two intervals intersect, they continue to grow, overlapping one another; in R, the pair of end points that meet at the intersection contribute no further to growth, and the number of intervals in R decreases by 1, whereas, in S^*, each end point of the pair continues to generate further chords according to independent Poisson processes of rate ρ, each of these then initiating further independent birth and growth processes.

This birth and growth process $S^*(n)$ agrees with R during the initial development, and only becomes substantially different when it has grown enough that there are many overlapping intervals. Its advantage is that it has a branching structure, and is thus much more easily analyzed. For instance, with initial conditions as above, let $S(n)$ denote the set of intervals generated up to time n starting from P and let $M(n)$ denote their number; let $S'(n)$ and $M'(n)$ denote the same quantities generated starting from P'. Then M and M' are *independent* pure birth chains with offspring distribution $1 + \mathrm{Po}\,(2\rho)$, and the centres of the intervals, excluding P and P', are independent and uniformly distributed on C. Hence $\mathbf{E}M(n) = (1 + 2\rho)^n$, and

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

forms a square integrable martingale, so that

$$(1 + 2\rho)^{-n}M(n) \;\to\; W_\rho \text{ a.s.} \tag{2.1}$$

for some W_ρ such that $W_\rho > 0$ a.s. and $\mathbf{E}W_\rho = 1$. Note that $\mathrm{Var}\,W(n) \le 1$ for all n. Similar formulae hold also for M', modified appropriately when the initial conditions are such that P' only initiates its birth and growth process starting at time 1, rather than at time 0; for example, we would then have $\mathbf{E}M'(n) = (1 + 2\rho)^{n-1}$ and $W'(n) := (1 + 2\rho)^{-(n-1)}M'(n) \to W'_\rho$ a.s.

Our strategy is to pick initial conditions as above, and to run the construction up to integer times τ_r, chosen in such a way that $R(n)$ and $S^*(n)$ are still (almost) the same for $n \le \tau_r$. We then use the probability that there is a pair of intervals, one in $S(\tau_r)$ and the other in $S'(\tau_r)$, that (i) intersect, but (ii) such that (roughly speaking) neither is contained in the other; this we can use as an approximation for the probability that there are points accessible from both P and P' in $R(\tau_r)$. Note that a condition of the form (ii) has to be included, because the smaller interval in any such pair would (probably) have been rejected in the R-construction; see the discussion around (2.5) for the details. This is the only way in which the distinction between R and S^* makes itself felt.

The time at which the first intersection occurs lies with high probability in an interval around the value

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

where $\lfloor x \rfloor$ denotes the largest integer no greater than x. This is because then $(1+2\rho)^{n_0} \approx \sqrt{L\rho}$, relevant for 'birthday problem' reasons to be explained below. It is in fact enough for the asymptotics to consider the interval of times $\tau_r = n_0 + r$, for r such that

$$|r| + 1 \;\leq\; \frac{1}{6\log(1+2\rho)} \log(L\rho)\,. \tag{2.2}$$

The choice of 6 in the denominator is such that, within this range, both $\eta_1(r,r')(L\rho)^{-\frac{1}{2}}$ and $\eta_3(r,r')(L\rho)^{-\frac{1}{4}}$, quantities that appear in the error bounds in Corollaries 3.4 and 3.6 and Theorem 3.8, are bounded (and mostly small); outside it, the intersection probability is close to either 0 or 1.

For an r satisfying (2.2) to exist, L and ρ must be such that $L\rho \geq (1+2\rho)^6$, a condition asymptotically satisfied if $L\rho \to \infty$ and ρ remains bounded, but imposing restrictions on the rate of growth of ρ as a function of L, if $\rho \to \infty$. In this choice, one already sees a typical manifestation of the effect of discretization. In the continuous model of Newman, Moore and Watts [13], either of L or ρ could have been arbitrarily fixed, and the other then determined by the value of the product $\frac{1}{2}L\rho$, the (expected) number of shortcuts, this being the only essential variable in their model. Here, in the discrete variant, the values of L and ρ each have their separate meaning in comparison with the value 1 of the discrete time step. In particular, if any such r exists, then it follows that

$$2 \;\leq\; \frac{\log(L\rho)}{3\log(1+2\rho)} \;\leq\; \tau_r \;\leq\; \frac{2\log(L\rho)}{3\log(1+2\rho)}; \tag{2.3}$$

we shall for convenience also always assume that $L\rho \geq 10$.

Let $\phi_0 := \phi_0(L,\rho)$ be defined by the equations

$$(1+2\rho)^{n_0} \;=\; \phi_0\sqrt{L\rho} \quad\text{ and }\quad (1+2\rho)^{-1} \;\leq\; \phi_0 \;\leq\; 1, \tag{2.4}$$

so that $\phi_0 = (L\rho)^{-1/2}(1+2\rho)^{n_0}$; note that $\phi_0 \approx 1$ if ρ is small. Then, writing $R_r = R(\tau_r)$, $S_r = S(\tau_r)$ and $M_r = M(\tau_r)$, we have $\mathbf{E}M_r = \phi_0\sqrt{L\rho}(1+2\rho)^r$; similarly, writing $r' = r$ if M' starts at time 0 and $r' = r-1$ if M' starts at time 1, we have $\mathbf{E}M'_r = \phi_0\sqrt{L\rho}(1+2\rho)^{r'}$. For later use, label the intervals in S_r as $I_1, \ldots, I_{M_r}$, and the intervals in S'_r as $J_1, \ldots, J_{M'_r}$, with the indices respecting chronological order: if $i < j$, then I_i was created earlier than or at the same time as I_j.

At least for small ρ, we can make a 'birthday problem' calculation. There are about $\phi_0^2 L\rho(1+2\rho)^{r+r'}$ pairs of intervals at time τ_r such that one is in S_r and the other in S'_r, and each is of typical length ρ^{-1}; thus the expected number of intersecting pairs of intervals is about

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

and this, in the chosen range of r, grows from almost nothing to the typically large value $2\phi_0^2(L\rho)^{1/3}$.

We now introduce the key event $A_{r,r'}$, that no interval in R_r can be reached from both P and P' — equivalent to $d(P,P') > 2(n_0+r)$ if $r' = r$, and to $d(P,P') > 2(n_0+r) - 1$ if $r' = r-1$ — whose probability we wish to approximate. We do so by way of an event defined in terms of S_r and S'_r, which we show to be almost the same. Clearly, if the intervals of S_r and S'_r have no intersection, then $A_{r,r'}$ is true. However, the event $\{I_i \cap J_j \neq \emptyset\}$ does not necessarily imply that $A_{r,r'}$ is not true, if either of I_i or J_j was 'descended' from one of the birth and growth processes initiated in S or S' independently of R, after a merging of intervals or the rejection of a chord, or if either I_i or J_j was itself centred at the second end–point q' of a rejected chord. As indicated above, the only one of these possibilities that has an appreciable effect is when the chord generating I_i is rejected because its q' belongs to the interval which grows to become J_j at time τ_r, or vice versa.

More precisely, we formulate our approximating event using not only knowledge of S_r and S'_r, but also that of the auxiliary random variables used to distinguish rejection or acceptance of a chord when both q and q' belong to $\partial R(n+1)$ for some n. We let $(Z_i,\, i \geq 1)$ and $(Z'_j,\, j \geq 1)$ be sequences of independent Bernoulli $\mathrm{Be}\,(\frac{1}{2})$–distributed random variables which are independent of each other and everything else, and if I_i is such that, for some j, $I_i \subset J_j$ but $I_i^1 \not\subset J_j$, where, for an interval $K = [a,b] \subset C$, $K^1 := [a-1, b+1]$, then we reject (accept) the (chord generating the) interval I_i in the construction of R if $Z_i = 0$ ($Z_i = 1$); the Z'_j are used similarly to randomize the acceptance or rejection of J_j. Note that I_i is always rejected if $I_i^1 \subset J_j$ for some j, and that I_i is always accepted if I_i is contained in none of the J_j. (Recall that $I_i^1 \subset J_j$ translates into the process starting from P' taking a shortcut into an interval that was already covered by the process starting from P).

Then we define

$$X_{ij} \;=\; \mathbf{1}\{I_i \cap J_j \neq \emptyset\}\mathbf{1}\{I_i^1 \not\subset J_j\}\mathbf{1}\{J_j^1 \not\subset I_i\}(1-K_{ij})(1-K'_{ji}), \tag{2.5}$$

where

$$K_{ij} \;:=\; \mathbf{1}\{I_i \subset J_j,\, Z_i = 0\}; \qquad K'_{ji} \;:=\; \mathbf{1}\{J_j \subset I_i,\, Z'_j = 0\},$$

so that X_{ij} certainly takes the value 1 if I_i and J_j are both realized in R and intersect one other, but may also take the value 1 if, for instance, I_i is not realized in R because $I_i^1 \subset J_{j'}$ for some $j' \neq j$.

We set

$$V_{r,r'} \;=\; \sum_{i=1}^{M_r} \sum_{j=1}^{M'_r} X_{ij}, \tag{2.6}$$

noting that therefore $\{V_{r,r'} = 0\} \subset A_{r,r'}$. In the next section, we show that these two events are in fact almost equal. Furthermore, we are able to approximate the distribution of $V_{r,r'}$ by a mixed Poisson distribution whose mixture distribution we can identify, and this gives us our approximation to $\mathbf{P}[A_{r,r'}]$.

After this preparation, we are in a position to summarize our main results. We let D denote the small worlds distance between a randomly chosen pair of points P and P' on C, so that, as above,

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

As approximation, we use the distribution of a random variable D^* whose distribution is given by

$$\mathbf{P}\left[D^* > 2n_0 + x\right] \;=\; \mathbf{E}\{e^{-\phi_0^2(1+2\rho)^x W_\rho W'_\rho}\}, \qquad x \in \mathbf{Z}, \tag{2.7}$$

where W_ρ and W'_ρ are independent copies of the random variable defined in (2.1). The following theorem (c.f. Theorem 3.10) gives conditions for the approximation to be good asymptotically. In this theorem, the parameter ρ may be chosen to vary with L, within rather broad limits. The derived quantities ϕ_0 and n_0 both implicitly depend on L and ρ, as does the distribution of D^*.

Theorem 2.1 *If $L\rho \to \infty$ and $\rho = \rho(L) = O(L^\beta)$, with $\beta < 4/31$, then*

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

The limiting behaviour for the two different $\rho(L)$-regimes corresponding to Corollary 3.11 and Theorem 3.16 is described in the following theorems. Once again, derived quantities such as ϕ_0, n_0, N_0 and x_0 all implicitly depend on L and ρ.

Theorem 2.2 *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*

$$\begin{aligned}\mathbf{P}\left[D^* \ge N_0+1\right] &\ge 1-(1+2\rho)^{-\alpha};\\ \mathbf{P}\left[D^* \ge N_0+2\right] &\le (1+2\rho)^{-1+\alpha}\log\{2(1+\rho)\},\end{aligned}$$

so that D^* concentrates almost all of its mass on N_0+1 if ρ is large, unless α is very close to 0 or 1.

Theorem 2.3 *If $\rho \to 0$, the distribution of $\rho(D^* - 2n_0)$ approaches that of the random variable T defined in [7], Corollary 3.10:*

$$\mathbf{P}[\rho(D^*-2n_0) > x] \;\to\; \mathbf{P}\left[T > x\right] \;=\; \int_0^\infty \frac{e^{-y}}{1+e^{2x}y}\,dy.$$

In the detailed versions below, the errors in these distributional approximations are also quantified.

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

$$\begin{aligned}\mathbf{P}[\rho(D^*-2n_0) > x] &= \mathbf{E}\{e^{-\phi_0^2(1+2\rho)^{x/\rho}W_\rho W'_\rho}\}\\ &\approx \mathbf{E}\{e^{-e^{2x}WW'}\} \;=\; \mathbf{P}[T > x],\end{aligned} \tag{2.8}$$

where W and W' are independent negative exponential NE(1) random variables, i.e. $\mathbf{P}(W > x) = e^{-x}$ for $x > 0$. Indeed, it follows from Lemma 3.14 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

$$\begin{aligned}\mathbf{P}[T > x \,|\, W, W'] &= e^{-e^{2x}WW'}\\ &= e^{-\exp\{2x+\log W+\log W'\}}\\ &= e^{-\exp\{2x-G_1-G_2\}},\end{aligned} \tag{2.9}$$

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

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

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

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

where G_1, G_2 and G_3 are independent random variables with the Gumbel distribution (see Janson [11] for an analogous result in a somewhat different context). The cumulants of T can thus immediately be deduced from those of the Gumbel distribution, given in Gumbel [9]:

$$\begin{aligned} \mathbf{E}T &= \gamma/2 \approx 0.2886; \\ \operatorname{Var} T &= \pi^2/8. \end{aligned}$$

Note that the conditional construction given above is visible in the arguments of the next section. The proofs are based on first conditioning on the processes S and S' up to time τ_r, and Corollary 3.6 below justifies an approximation of the same form as in (2.8), with W and W' replaced by W_ρ and W'_ρ. These random variables are, however, essentially determined by the early stages of the respective pure birth processes; the extra randomness, once the values of W_ρ and W'_ρ have been fixed, comes from the random arrangement of the intervals of S_r and S'_r 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

3.1 Poisson approximation for the intersection probability

The first step in the argument outlined above is to establish a Poisson approximation theorem for the distribution of $V_{r,r'}$, conditional on $\mathcal{F}_r := \sigma(M(l), M'(l),\, 0 \le l \le \tau_r)$. This is rather easy, because $V_{r,r'}$ is a sum of dissociated random variables.

Proposition 3.1 *Let $V_{r,r'}$ be defined as in (2.6). Then, if P and P' are chosen uniformly and independently on C, we have*

$$d_{TV}\{\mathcal{L}(V_{r,r'} \,|\, \mathcal{F}_r), \operatorname{Po}(\lambda_r(M, M'))\} \le 8(M_r + M'_r)\tau_r/L,$$

where

$$\lambda_r(M, M') := \sum_{i=1}^{M_r} \sum_{j=1}^{M'_r} \mathbf{E}\{X_{ij} \,|\, \mathcal{F}_r\}.$$

Proof: Since, at time τ_r, each interval has length at most $2\tau_r$, and because their centres are independently and uniformly distributed on C, it follows that

$$\begin{aligned} &\mathbf{E}\{X_{ij} \,|\, \mathcal{F}_r\} \le 4\tau_r L^{-1}, \quad \mathbf{E}\{X_{ij}X_{il} \,|\, \mathcal{F}_r\} \le 4\tau_r L^{-1}\mathbf{E}\{X_{ij} \,|\, \mathcal{F}_r\} \\ &\text{and} \quad \mathbf{E}\{X_{ij}X_{kj} \,|\, \mathcal{F}_r\} \le 4\tau_r L^{-1}\mathbf{E}\{X_{ij} \,|\, \mathcal{F}_r\}, \end{aligned}$$

for all $1 \leq i, k \leq M_r$ and $1 \leq j, l \leq M'_r$, and that X_{ij} is independent of the random variables $\{X_{kl},\, 1 \leq k \neq i \leq M_r,\, 1 \leq l \neq j \leq M'_r\}$. The proposition now follows directly from the Stein–Chen local approach: see ([6], Theorem 1.A). □

Remark. If P and P' are not chosen at random, but are fixed points of C, the result of Proposition 3.1 remains essentially unchanged, provided that P and P' are more than an arc distance of $2n_0 + r + r'$ apart. The only difference is that then $X_{11} = 0$ a.s., and that $\lambda_r(M, M')$ is replaced by $\lambda_r(M, M') - 2(2n_0 + r + r')L^{-1}$, because the two points P and P' are already forced to be at least $2n_0 + r + r'$ apart in both directions on the circle. If P and P' are less than $2n_0 + r + r'$ apart, then $\mathbf{P}[A_{r,r'}] = 0$.

From Proposition 3.1, $\mathbf{P}[V_{r,r'} = 0]$ can be well approximated by $\mathbf{E}\{e^{-\lambda_r(M,M')}\}$, provided that we can show that $\tau_r \mathbf{E} M_r/L$ is small. The next lemma shows that, with somewhat better control of the distribution of M_r, the probability $\mathbf{P}[A_{r,r'}]$ that we are interested in is itself well approximated by $\mathbf{P}[V_{r,r'} = 0]$. To prove it, we compare suitably chosen random variables in the joint construction.

Lemma 3.2 *With notation as above, we have*

$$\begin{aligned} 0 &\leq \mathbf{P}[A_{r,r'}] - \mathbf{P}[V_{r,r'} = 0] \\ &\leq 16\tau_r^2 L^{-2} \mathbf{E}\{M_r M'_r (M_r + M'_r)(1 + \log M_r + \log M'_r)\}. \end{aligned}$$

Proof: The proof requires some extra notation, involving ancestry. Each interval I in S is created at some time n by a single label 2 endpoint q' of a chord $\{q, q'\}$, whose label 1 endpoint belonged to the extension of an interval of $S(n-1)$. This latter interval is considered to be the parent of I. Using this notion, we can construct a 'family tree' for the intervals of S. In particular, for any $l \geq 1$, we can take the set of intervals $\mathbb{I}_l := \{I_1, \ldots, I_l\}$ to be 'ancestor' intervals, and then determine, for each i, the index $A^l(i)$, $1 \leq A^l(i) \leq l$, of the most recent ancestor of I_i among $\mathbb{I}_l$. Then, letting $S_r^i := \bigcup_{l \neq i} I_l$ and $S_r'^j := \bigcup_{l \neq j} J_l$, we define

$$\begin{aligned} H_{i1} &:= \{I_i \cap S_r^i \neq \emptyset\}; \qquad H_{i2} := \bigcup_{1 \leq l < i} \left(\{I_l \cap S_r^l \neq \emptyset\} \cap \{A^l(i) = l\}\right) \\ H_{i3} &:= \bigcup_{1 \leq l < i} \left(\{I_l \cap S'_r \neq \emptyset\} \cap \{A^l(i) = l\}\right), \end{aligned}$$

and set $H_i := \bigcup_{v=1}^{3} H_{iv}$; we also define H'_{iv}, $1 \leq v \leq 3$ and H'_i analogously. Thus H_{i1} is the event that the interval I_i is not isolated in S_r; H_{i2} is the event that some ancestor of interval I_i is not isolated in S_r; and H_{i3} is the event that some ancestor of interval I_i intersects S'_r; H_i summarises the event that I_i or one or more of its ancestors intersect either S'_r or the remainder of S_r.

Then, with X_{ij} defined as in (2.5),

$$\{V_{r,r'} = 0\} \subset A_{r,r'} \subset \{\widetilde{V}_{r,r'} = 0\}, \tag{3.1}$$

where

$$\widetilde{V}_{r,r'} := \sum_{i=1}^{M_r} \sum_{j=1}^{M'_r} X_{ij} I[H_i^c]\, I[H_j'^c],$$

so that

$$\begin{aligned}\mathbf{P}[A_{r,r'} \setminus \{V_{r,r'}=0\}] \;&\le\; \mathbf{P}[V_{r,r'} \neq \widetilde{V}_{r,r'}] \;\le\; \mathbf{E}(V_{r,r'} - \widetilde{V}_{r,r'}) \\ &\le\; \mathbf{E}\left\{ \sum_{i=1}^{M_r} \sum_{j=1}^{M'_r} X_{ij} \sum_{v=1}^{3} (I[H_{iv}] + I[H'_{jv}]) \right\}.\end{aligned} \tag{3.2}$$

Now, conditional on $\mathcal{F}_r$, the indicator X_{ij} is (pairwise) independent of each of the events H_{iv} and H'_{jv}, $1 \le v \le 3$, because H_{i1}, H_{i2} and H'_{j3} are each independent of ζ'_j, the centre of J_j, and H'_{j1}, H'_{j2} and H_{i3} are each independent of ζ_i. Moreover, the event $\{A^l(i) = l\}$ is independent of M, M' and all ζ_i's and ζ'_j's, and has probability $1/l$, since the l birth processes, one generated from each interval I_v, $1 \le v \le l$, which combine to make up S from time that I_l was initiated onwards, are independent and identically distributed. Hence, observing also that no interval at time τ_r can have length greater than $2\tau_r$, it follows that

$$\begin{aligned}\mathbf{E}\{X_{ij} I[H_{i1}] \,|\, \mathcal{F}_r\} \;&\le\; 4\tau_r L^{-1}(M_r - 1)\mathbf{E}\{X_{ij} \,|\, \mathcal{F}_r\}, \\ \mathbf{E}\{X_{ij} I[H_{i2}] \,|\, \mathcal{F}_r\} \;&\le\; 4\tau_r L^{-1}(M_r - 1) \sum_{l=1}^{i-1} l^{-1} \mathbf{E}\{X_{ij} \,|\, \mathcal{F}_r\} \\ &\le\; 4\tau_r L^{-1}(M_r - 1) \log M_r \mathbf{E}\{X_{ij} \,|\, \mathcal{F}_r\}\end{aligned}$$

and

$$\mathbf{E}\{X_{ij} I[H_{i3}] \,|\, \mathcal{F}_r\} \;\le\; 4\tau_r L^{-1} M'_r \log M_r \mathbf{E}\{X_{ij} \,|\, \mathcal{F}_r\}.$$

Hence it follows that

$$\begin{aligned}&\sum_{i=1}^{M_r} \sum_{j=1}^{M'_r} \sum_{v=1}^{3} \mathbf{E}\{X_{ij} I[H_{iv}] \,|\, \mathcal{F}_r\} \\ &\qquad \le\; 4\tau_r L^{-1}\{(M_r - 1)(1 + \log M_r) + M'_r \log M_r\} \lambda_r(M, M'),\end{aligned}$$

and combining this with a similar contribution from the quantities $\mathbf{E}\{X_{ij} I[H'_{jv}] \,|\, \mathcal{F}_r\}$ gives

$$\begin{aligned}&\mathbf{E}\left\{ \sum_{i=1}^{M_r} \sum_{j=1}^{M'_r} X_{ij} \sum_{v=1}^{3} (I[H_{iv}] + I[H'_{jv}]) \,\Big|\, \mathcal{F}_r \right\} \\ &\qquad \le\; 4\tau_r L^{-1}(M_r + M'_r - 1)(1 + \log M_r + \log M'_r) \lambda_r(M, M').\end{aligned}$$

Observing that $\lambda_r(M, M') \le 4\tau_r L^{-1} M_r M'_r$ completes the proof. □

To apply Proposition 3.1 and Lemma 3.2, we need in particular to bound the moments of M_r appearing in Lemma 3.2, which we accomplish in the following lemma. The corresponding results for M'_r can be obtained by replacing r with r'.

Lemma 3.3 *The random variable $M(n)$ has as probability generating function*

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

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

$$\begin{aligned}\mathbf{E}M_r &= \phi_0\sqrt{L\rho}(1+2\rho)^r, \quad \mathbf{E}M_r^2 \;\le\; 2\phi_0^2 L\rho(1+2\rho)^{2r},\\ \mathbf{E}M_r^3 &\le 6\phi_0^3(L\rho)^{3/2}(1+2\rho)^{3r}\end{aligned}$$

and

$$\begin{aligned}\mathbf{E}\{M_r\log M_r\} &\le \{\tau_r\log(1+2\rho)+2\}\phi_0\sqrt{L\rho}(1+2\rho)^r,\\ \mathbf{E}\{M_r^2\log M_r\} &\le 2\{\tau_r\log(1+2\rho)+3\}\phi_0^2 L\rho(1+2\rho)^{2r}.\end{aligned}$$

Proof: Since $M(n)$ is a branching process with $1+\mathrm{Po}\,(2\rho)$ offspring distribution, the probability generating function is immediate, as is the formula for its first moment $m_n := m_n^{(1)} = (1+2\rho)^n$, where we write $m_n^{(v)} := \mathbf{E}\{M^v(n)\}$. Considering the outcome of the first generation, and letting Z denote a Poisson random variable with mean $1+2\rho$, we then have the recurrences

$$\begin{aligned}m_n^{(2)} &= m_{n-1}^{(2)}\mathbf{E}Z + m_{n-1}^2\mathbf{E}\{Z(Z-1)\}\\ &= (1+2\rho)m_{n-1}^{(2)} + 4\rho(1+\rho)(1+2\rho)^{2n-2}\end{aligned}$$

and

$$\begin{aligned}m_n^{(3)} &= m_{n-1}^{(3)}\mathbf{E}Z + 3m_{n-1}m_{n-1}^{(2)}\mathbf{E}\{Z(Z-1)\} + m_{n-1}^3\mathbf{E}\{Z(Z-1)(Z-2)\}\\ &= (1+2\rho)m_{n-1}^{(3)} + 12\rho(1+\rho)(1+2\rho)^{n-1}m_{n-1}^{(2)} + 4\rho^2(3+2\rho)(1+2\rho)^{3n-3}.\end{aligned}$$

From the first recurrence, it follows easily that

$$m_n^{(2)} \;\le\; 2(1+\rho)(1+2\rho)^{2n-1} \;\le\; 2(1+2\rho)^{2n}, \tag{3.3}$$

and then from the second that

$$\begin{aligned}m_n^{(3)} &\le \left\{24\rho(1+\rho)^2 + 4\rho^2(3+2\rho)(1+2\rho)\right\}(1+2\rho)^{3n-4} \,/\, \left\{1-(1+2\rho)^{-2}\right\}\\ &\le 6(1+2\rho)^{3n}.\end{aligned}$$

The bounds for the moments of M_r follow by replacing n by τ_r. The remaining bounds are deduced using the inequality

$$\log m \;\le\; n\log(1+2\rho) + m(1+2\rho)^{-n}. \qquad \square$$

These estimates can be directly applied in Lemma 3.2 and Proposition 3.1. Define

$$\eta_1(r,r') \;:=\; 704\rho^2(n_0+r)^3\log(1+2\rho)\phi_0^3(1+2\rho)^{2r+r'}; \tag{3.4}$$

$$\eta_2(r,r') \;:=\; 16\rho(n_0+r)\phi_0(1+2\rho)^r, \tag{3.5}$$

and note, for the calculations, that $3\tau_r\log(1+2\rho)\ge 2$, from (2.3) and because $L\rho\ge 10$.

Corollary 3.4 *We have*

$$0 \;\le\; \mathbf{P}[A_{r,r'}] - \mathbf{P}[V_{r,r'}=0] \;\le\; \eta_1(r,r')(L\rho)^{-1/2},$$

and hence, recalling that $A_{r,r'} = \{D > 2n_0+r+r'\}$,

$$|\mathbf{P}[D>2n_0+r+r'] - \mathbf{E}\exp\{-\lambda_r(M,M')\}| \;\le\; \{\eta_1(r,r')+\eta_2(r,r')\}(L\rho)^{-1/2}. \qquad \square$$

3.2 The shortest path length: first asymptotics

In view of Corollary 3.4, the asymptotics of $\mathbf{P}[D > 2n_0 + r + r']$, where D denotes the "small world" shortest path distance between P and P', require closer consideration of the quantity $\mathbf{E}(\exp\{-\lambda_r(M, M')\})$. Now $\lambda_r(M, M')$ can be expressed quite neatly in terms of the processes M and M'. Such a formula is given in following lemma, together with a second, less elegant representation, which is useful for the asymptotics. For a sequence $H(\cdot)$, we use the notation $\nabla H(s)$ for the backward difference $H(s) - H(s-1)$. We define $X(s) := M(s) - (1+2\rho)M(s-1)$ and $X'(s) := M'(s) - (1+2\rho)M'(s-1)$, and set $M(-1) = M'(-1) = 0$; we also write

$$\begin{aligned} E_r(M, M') \;:=\; & \sum_{s=2}^{\tau_r} (2\rho\{M(s-1)X'(s) + M'(s-1)X(s)\} + X(s)X'(s)) \\ & \qquad - 3M(0)M'(0) + \nabla M(1)\nabla M'(1). \end{aligned} \tag{3.6}$$

Lemma 3.5 *With the above definitions, we have*

$$\begin{aligned} L\lambda_r(M, M') &= 4\sum_{s=0}^{\tau_r - 1} M(s)M'(s) + M_r M'_r - \sum_{s=0}^{\tau_r} \nabla M(s)\nabla M'(s) \\ &= 4(1-\rho^2)\sum_{s=1}^{\tau_r - 1} M(s)M'(s) + M_r M'_r - E_r(M, M'). \end{aligned}$$

Proof: Let the intervals I_i and J_j have lengths s_i and u_j respectively. By construction, both lengths must be even integers, and the centres of the intervals are independently and uniformly chosen on C. Then, fixing I_i, and supposing that $u_j \le s_i - 2$, there are two intervals on C, each of length u_j, in whose union the centre of J_j must lie, if I_i and J_j are to overlap without J_j being contained in I_i; and there are two further intervals, each of length 1, for which one has $J_j \subset I_i$ but $J_j^1 \not\subset I_i$, in which case $\mathbf{E}(1 - K'_{ji}) = \frac{1}{2}$. From these and similar considerations, it follows from (2.5) that

$$\mathbf{E}\{X_{ij} \,|\, \mathcal{F}_r\} \;=\; L^{-1}\{2\min(s_i, u_j) + 1 - \delta_{s_i,u_j}\} \;=:\; L^{-1} f(s_i, u_j),$$

where δ_{kl} denotes the Kronecker delta and $f(x,y) := 2\min(x,y) + 1 - \delta_{x,y}$. Hence

$$\begin{aligned} L\lambda_r(M, M') &= \sum_{l=0}^{\tau_r}\sum_{l'=0}^{\tau_r} \nabla M(l)\nabla M'(l') f(2(\tau_r - l), 2(\tau_r - l')) \\ &= \sum_{l=0}^{\tau_r} \Big\{ (4(\tau_r - l) + 1)\{\nabla M(l) M'(l-1) + \nabla M'(l) M(l-1)\} \\ &\qquad\qquad + 4(\tau_r - l)\nabla M(l)\nabla M'(l) \Big\} \\ &= \sum_{l=0}^{\tau_r} \Big\{ (4(\tau_r - l) + 1)\nabla\{M(l)M'(l)\} - \nabla M(l)\nabla M'(l) \Big\}, \end{aligned}$$

and summation by parts completes the proof of the first formula. The second follows from the observation that $\nabla M(s) = X(s) + 2\rho M(s-1)$. $\square$

As a result of this lemma, combined with Corollary 3.4, we can approximate the distribution of the shortest path length D in terms of that of the random variable D^*, introduced in (2.7).

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

$$|\mathbf{P}[D > s] - \mathbf{P}[D^* > s]| \ \to \ 0$$

uniformly in $|s - 2n_0| + 2 \le \frac{2\log(L\rho)}{7\log(1+2\rho)}$.

Proof: For s in the given range, we write $r = \lceil (s - 2n_0)/2 \rceil$ and $r' = s - 2n_0 - r$, and observe that, under the stated conditions, τ_r tends to infinity as $L \to \infty$ at least as fast at $c\log(L\rho)$, for some $c > 0$, because of (2.3). We also observe that

$$\{\eta_1(r, r') + \eta_2(r, r')\}(L\rho)^{-1/2} \ \to \ 0$$

uniformly in the given range of r.

Now $W(n) = (1+2\rho)^{-n}M(n) \to W_\rho$ a.s. and, for $X(s)$ appearing in the definition (3.6) of $E_r(M, M')$, we have $\mathbf{E}\{X(s) \,|\, M(0), \ldots, M(s-1)\} = 0$ and

$$\mathrm{Var}\, X(s) \ = \ 2\rho\mathbf{E}M(s-1) \ = \ 2\rho(1+2\rho)^{s-1}; \tag{3.7}$$

furthermore, $(1+2\rho)^{2n_0+r+r'} = \phi_0^2(1+2\rho)^{r+r'}L\rho$. Hence, and from Lemma 3.5, it follows that

$$\begin{aligned}
\lambda_r(M, M') \ &\sim \ (1+2\rho)^{2n_0+r+r'}L^{-1}W_\rho W'_\rho \left\{ 4(1-\rho^2)\sum_{l=1}^{\tau_r - 1}(1+2\rho)^{-2l} + 1 \right\} \\
&\sim \ \phi_0^2 L\rho(1+2\rho)^{r+r'}L^{-1}\left\{ \frac{4(1-\rho^2)}{4\rho(1+\rho)} + 1 \right\} W_\rho W'_\rho \\
&= \ \phi_0^2(1+2\rho)^{r+r'}W_\rho W'_\rho,
\end{aligned} \tag{3.8}$$

uniformly for r in the given range. This, together with the fact that

$$\mathbf{P}[D^* > 2n_0 + r + r'] \ = \ \mathbf{E}\{e^{-\phi_0^2(1+2\rho)^x W_\rho W'_\rho}\},$$

from (2.7), completes the proof. □

Hence $\mathbf{P}[D > s]$ can be approximated in terms of the distribution of D^*, which is itself determined by that 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 the negative exponential distribution NE (1) with mean 1, but genuinely depends on ρ. Its properties are not so easy to derive, though moments can be calculated, and, in particular,

$$\mathbf{E}W_\rho \ = \ 1; \qquad \mathrm{Var}\, W_\rho \ = \ 1/(1+2\rho); \tag{3.9}$$

it is also shown in Lemma 3.14 that $\mathcal{L}(W_\rho)$ is close to NE (1) for ρ small.

3.3 The shortest path length: error bounds

The simple asymptotics of Corollary 3.6 can be sharpened, to provide bounds for the errors involved. The main effort lies in proving a more accurate approximation to $\lambda_r(M, M')$ than that given in (3.8).

At first sight surprisingly, it turns out that it is not actually necessary for the time τ_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.

Lemma 3.7 *We have*

$$\lambda_r(M, M') \;=\; (1+2\rho)^{r+r'}\phi_0^2\{W(\tau_r)W'(\tau_r) + U_r\},$$

where U_r, given in (3.11) below, is such that

$$\mathbf{E}|U_r| \;\le\; 11(1+2\rho)^{(3-\tau_r)/2}.$$

Proof: We begin by observing that

$$\begin{aligned}
&\rho(1+2\rho)^{-(2n_0+r+r')}\sum_{s=1}^{\tau_r-1} M(s)M'(s)\\
&\quad= \rho\sum_{l=1}^{\tau_r-1}(1+2\rho)^{-2l}W(\tau_r-l)W'(\tau_r-l)\\
&\quad= \frac{1}{4(1+\rho)}W(\tau_r)W'(\tau_r) - U_{r1} + U_{r2},
\end{aligned}\tag{3.10}$$

where

$$\begin{aligned}
U_{r1} &= \frac{W(\tau_r)W'(\tau_r)}{4(1+\rho)(1+2\rho)^{2(\tau_r-1)}},\\
U_{r2} &= \rho\sum_{l=1}^{\tau_r-1}(1+2\rho)^{-2l}W_r^*(l)
\end{aligned}$$

and

$$W_r^*(l) \;:=\; W(\tau_r-l)W'(\tau_r-l) - W(\tau_r)W'(\tau_r).$$

Hence, since also

$$\rho(1+2\rho)^{-(2n_0+r+r')}M_rM_r' \;=\; \rho W(\tau_r)W'(\tau_r),$$

we have shown that

$$\begin{aligned}
&L\rho(1+2\rho)^{-(2n_0+r+r')}\lambda_r(M, M')\\
&\quad= W(\tau_r)W'(\tau_r) + 4(1-\rho^2)(U_{r2}-U_{r1}) - U_{r3},
\end{aligned}$$

where

$$U_{r3} \;:=\; \rho(1+2\rho)^{-(2n_0+r+r')}E_r(M, M') \;=\; \sum_{v=1}^{4} V_{rv},$$

with

$$\begin{aligned}
V_{r1} &= 2\rho^2 \sum_{s=2}^{\tau_r} (1+2\rho)^{-2(\tau_r - s)-1} W(s-1)\nabla W'(s), \\
V_{r2} &= 2\rho^2 \sum_{s=2}^{\tau_r} (1+2\rho)^{-2(\tau_r - s)-1} \nabla W(s) W'(s-1), \\
V_{r3} &= \rho \sum_{s=2}^{\tau_r} (1+2\rho)^{-2(\tau_r - s)} \nabla W(s)\nabla W'(s)
\end{aligned}$$

and

$$V_{r4} = \rho(1+2\rho)^{-(2n_0+r+r')}\{\nabla M(1)\nabla M'(1) - 3\delta_{r,r'}\}.$$

Thus we have assembled U_r; recalling that $(1+2\rho)^{2n_0} = \phi_0^2 L\rho$ from (2.4), we have

$$U_r = 4(1-\rho^2)(U_{r2} - U_{r1}) - U_{r3}. \tag{3.11}$$

It remains to bound $\mathbf{E}|U_r|$. It is immediate that $U_{r1} \ge 0$, so that

$$\mathbf{E}|U_{r1}| = \frac{1}{4(1+\rho)(1+2\rho)^{2(\tau_r - 1)}}. \tag{3.12}$$

Then $\mathbf{E}|U_{r2}| \le \sqrt{\mathbf{E}U_{r2}^2}$; to bound the latter, we begin by defining

$$v_{lm} := \mathbf{E}\{W_r^*(l)W_r^*(m)\}, \quad 1 \le l \le m \le \tau_r - 1,$$

and $v_{lm} = v_{ml}$ if $l > m$, so that

$$\mathbf{E}U_{r2}^2 = \rho^2 \sum_{l=1}^{\tau_r - 1}\sum_{m=1}^{\tau_r - 1} (1+2\rho)^{-2(l+m)} v_{lm}.$$

As W and W' are independent martingales, with $\nabla W(s) = (1+2\rho)^{-s}X(s)$, we have from (3.3) and (3.7) that

$$\mathbf{E}W^2(s) \le 2, \ s \ge 1; \qquad \mathbf{E}\{(W(s) - W(t))^2\} \le (1+2\rho)^{-(s+1)}, \ 1 \le s \le t \le \infty. \tag{3.13}$$

Thus, for $1 \le l \le m \le \tau_r - 1$,

$$\begin{aligned}
v_{lm} &= \mathbf{E}\{(W_r^*(l))^2\} \\
&= \mathbf{E}\{(W'(\tau_r - l)^2\}\mathbf{E}\{(W(\tau_r - l) - W(\tau_r))^2\} \\
&\qquad + \mathbf{E}\{W^2(\tau_r)\}\mathbf{E}\{(W'(\tau_r - l) - W'(\tau_r))^2\} \\
&\le 4(1+2\rho)^{-(\tau_r - l + 1)}.
\end{aligned}$$

Hence it follows that

$$\begin{aligned}
\mathbf{E}U_{r2}^2 &\le 2\rho^2 \sum_{l=1}^{\tau_r - 1} \frac{(1+2\rho)^{-4l}}{4\rho(1+\rho)}\, 4(1+2\rho)^{-(\tau_r - l+1)} \\
&\quad + \frac{4\rho^2}{(1+2\rho)^3 - 1}\,(1+2\rho)^{-\tau_r - 1} \\
&\le \left\{\frac{2}{\rho(1+\rho)} + 4\right\} \frac{\rho^2}{(1+2\rho)^3 - 1}\,(1+2\rho)^{-\tau_r - 1} \\
&\le (1+\rho)^{-1}(1+2\rho)^{-\tau_r - 1},
\end{aligned}$$

and so

$$\mathbf{E}|U_{r2}| \;\le\; (1+\rho)^{-1}(1+2\rho)^{-\tau_r/2}. \tag{3.14}$$

In order to bound $\mathbf{E}|U_{r3}|$, we begin by noting that $\mathbf{E}V_{r1} = \mathbf{E}V_{r2} = \mathbf{E}V_{r3} = 0$. Then, by the orthogonality of the martingale differences $\nabla W(s)$ and $\nabla W'(s)$, we have

$$\begin{aligned}
\mathbf{E}V_{r1}^2 &= 4\rho^4 \sum_{s=2}^{\tau_r} (1+2\rho)^{-4(\tau_r-s)-2}\mathbf{E}W^2(s-1)\mathbf{E}\{(\nabla W'(s))^2\} \\
&\le 16\rho^5 \sum_{l=0}^{\tau_r-2} (1+2\rho)^{-(\tau_r+3+3l)} \\
&\le \frac{8\rho^4(1+2\rho)}{1+2\rho+4\rho^2}\,(1+2\rho)^{-\tau_r} \;\le\; 8\rho^4(1+2\rho)^{-\tau_r-1},
\end{aligned}$$

and the same bound is true for $\mathbf{E}V_{r2}^2$ as well. Then, similarly, we obtain the bound

$$\begin{aligned}
\mathbf{E}V_{r3}^2 &\le \rho^2 \sum_{s=2}^{\tau_r} (1+2\rho)^{-4(\tau_r-s)}\, 4\rho^2(1+2\rho)^{-2s-1} \\
&\le \frac{\rho^3(1+2\rho)}{1+\rho}\,(1+2\rho)^{-2\tau_r} \;\le\; 2\rho^3(1+2\rho)^{-2\tau_r},
\end{aligned}$$

and, finally,

$$\mathbf{E}|V_{r4}| \;\le\; \rho(1+2\rho)^{-2\tau_r+1}\max\{4\rho^2, 3\}.$$

Hence

$$\begin{aligned}
\mathbf{E}|U_{r3}| \;\le\;& 4\sqrt{2}\rho^2(1+2\rho)^{-(\tau_r+1)/2} \\
&+ \sqrt{2}\rho^{3/2}(1+2\rho)^{-\tau_r} + 4\rho\max\{1,\rho^2\}(1+2\rho)^{-2\tau_r+1}.
\end{aligned} \tag{3.15}$$

Substituting from (3.12), (3.14) and (3.15) into (3.11), we obtain, after some calculation,

$$\mathbf{E}|U_r| \;\le\; 11\max\{1,\rho^{3/2}\}(1+2\rho)^{-\tau_r/2},$$

proving the result. □

From Lemma 3.7, observing that $|e^{-x}-e^{-y}| \le |x-y|$ for $x, y \ge 0$, we note that

$$\begin{aligned}
&\Big|\mathbf{E}\exp\{-\lambda_r(M,M')\} - \mathbf{E}\exp\{-\phi_0^2(1+2\rho)^{r+r'}W(\tau_r)W'(\tau_r)\}\Big| \\
&\quad\le\; \phi_0^2(1+2\rho)^{r+r'}\mathbf{E}|U_r|
\end{aligned} \tag{3.16}$$

and, because also $\mathbf{E}W(n) = 1$ for all n, that

$$\begin{aligned}
&\Big|\mathbf{E}\exp\{-\phi_0^2(1+2\rho)^{r+r'}W(\tau_r)W'(\tau_r)\} - \mathbf{E}\exp\{-\phi_0^2(1+2\rho)^{r+r'}W_\rho W'_\rho\}\Big| \\
&\quad\le\; \phi_0^2(1+2\rho)^{r+r'}\{\mathbf{E}|W_\rho - W(\tau_r)| + \mathbf{E}|W'_\rho - W'(\tau_r)|\}.
\end{aligned} \tag{3.17}$$

Using these results, we obtain the following theorem, giving a pointwise bound for the difference between the distribution functions of D and D^*.

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

$$\begin{aligned}&|\mathbf{P}[D > 2n_0 + r + r'] - \mathbf{P}[D^* > 2n_0 + r + r']| \\ &\quad \le \; \{\eta_1(r,r') + \eta_2(r,r')\}(L\rho)^{-1/2} + \eta_3(r,r')(L\rho)^{-1/4},\end{aligned}$$

where η_1, η_2 are given in (3.4) and (3.5),

$$\eta_3(r,r') \; := \; 13\phi_0^{3/2}(1+2\rho)^{r'+(r+3)/2}, \tag{3.18}$$

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

Proof: We use Corollary 3.4, Lemma 3.7 and (3.16) and (3.17) to give

$$\begin{aligned}&\left|\mathbf{P}[D > 2n_0 + r + r'] - \mathbf{E}\{e^{-\phi_0^2(1+2\rho)^{r+r'}W_\rho W'_\rho}\}\right| \\ &\quad \le \; \{\eta_1(r,r') + \eta_2(r,r')\}(L\rho)^{-1/2} + \phi_0^2(1+2\rho)^{r+r'}\mathbf{E}|U_r| \\ &\qquad + \phi_0^2(1+2\rho)^{r+r'}\{\mathbf{E}|W_\rho - W(\tau_r)| + \mathbf{E}|W'_\rho - W'(\tau_r)|\},\end{aligned} \tag{3.19}$$

where $\mathbf{E}|U_r| \le 11(1+2\rho)^{(3-\tau_r)/2}$. Now

$$\mathbf{E}\{(W'_\rho - W'(\tau_r))^2\} \; \le \; (1+2\rho)^{-\tau_r}$$

from (3.13) (note that, if $r' = r-1$, W' is a step behind W), and hence

$$\mathbf{E}|W_\rho - W(\tau_r)| + \mathbf{E}|W'_\rho - W'(\tau_r)| \; \le \; 2(1+2\rho)^{-\tau_r/2}.$$

The theorem now follows from (2.7). $\square$

For fixed ρ and r, r' bounded, the error in Theorem 3.8 is of order $(L\rho)^{-1/4}$. However, if $r + r'$ grows, it rapidly becomes larger, because of the powers of $(1+2\rho)$ appearing in the quantities $\eta_l(r,r')$. Thus, in order to translate it into a uniform distributional approximation, a separate bound for the upper tail of $\mathcal{L}(D^*)$ is needed. This is given by way of the following lemmas.

Lemma 3.9 *If X is a random variable such that*

$$\mathbf{E}\{e^{-\theta X}\} \le (1+\theta)^{-1}, \tag{3.20}$$

then, for independent copies X_1, X_2 of X, we have

$$\mathbf{E}\left(e^{-\theta X_1 X_2}\right) \le \theta^{-1}\log(1+\theta).$$

In particular, for all $\theta, \rho > 0$,

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

Proof: We begin by noting that

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

Hence, applying (3.20) twice, and because the function $(1+t)^{-1}$ is decreasing in $t \geq 0$, we obtain

$$\begin{aligned}
\mathbf{E}\left(e^{-\theta X_1 X_2}\right) &\leq \mathbf{E}\{(1+\theta X_1)^{-1}\} \\
&= \theta^{-1}\int_0^\infty \mathbf{E}e^{-tX_1}e^{-t/\theta}\,dt \\
&\leq \theta^{-1}\int_0^\infty (1+t)^{-1}e^{-t/\theta}\,dt \\
&\leq \theta^{-1}\int_0^\theta (1+t)^{-1}\,dt \;=\; \theta^{-1}\log(1+\theta),
\end{aligned}$$

as required.

Now, the offspring generating function of the birth process M satisfies

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

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

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

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. So (3.20) holds, and the first part of the lemma applies, giving the assertion. □

We can now prove the following uniform bound on the distance between the distributions of D and D^*. For probability distributions Q and Q' on $\mathbf{R}$, we use d_K to denote the Kolmogorov distance:

$$d_K(Q,Q') \;:=\; \sup_x |Q\{(-\infty,x]\} - Q'\{(-\infty,x]\}|.$$

Theorem 3.10 *For D the shortest path between randomly chosen points P and P' and D^* with distribution as in (2.7), we have*

$$\begin{aligned}
&d_K(\mathcal{L}(D),\mathcal{L}(D^*)) \\
&\quad = O\left(\log(L\rho)(1+2\rho)^{3/2}(L\rho)^{-\frac{1}{7}} + \left(\frac{\rho}{\log(1+2\rho)}\right)^2 (1+2\rho)^{1/2}(L\rho)^{-\frac{2}{7}}\log^3(L\rho)\right).
\end{aligned}$$

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

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

Proof: From Lemma 3.9 and (2.7), it follows that

$$\mathbf{P}[D^* > 2n_0 + x] \;\le\; \phi_0^{-2}(1+2\rho)^{-x}(1 + x\log(1+2\rho)) \tag{3.21}$$

for any $x > 0$, since $\phi_0 \le 1$ and $\log(1+y) \le 1 + \log y$ 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

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

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 the bound (3.21) evaluated at the value $x = \lceil \frac{\frac{1}{7}\log(L\rho) - 2\log\phi_0}{\log(1+2\rho)} \rceil$, which gives rise to a term of order $O\left((L\rho)^{-\frac{1}{7}}\log(L\rho)\right)$, 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 3.10, it is clear that $\mathcal{L}(D)$ becomes concentrated on very few values, once ρ is large, since the factor $\phi_0^2(1+2\rho)^x$ in the exponent in (2.7) is multiplied by the large factor $(1+2\rho)$ if x is increased by 1. The following corollary makes this more precise.

Corollary 3.11 *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

$$\mathbf{P}[D^* \ge N_0 + 1] \ge 1 - (1+2\rho)^{-\alpha},$$

and

$$\mathbf{P}[D^* \ge N_0 + 2] \;=\; \mathbf{E}\exp\{-(1+2\rho)^{1-\alpha}W_\rho W_\rho'\} \;\le\; (1+2\rho)^{-1+\alpha}\log\{2(1+\rho)\}.$$

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

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

as $\mathbf{E}W_\rho = 1$, and from Lemma 3.9 with $\theta = (1+2\rho)^{1-\alpha}$. $\square$

Thus the distribution of D^* is essentially concentrated on the single value $N_0 + 1$ if ρ is large and α is bounded away from 0 and 1. If, for instance, α is close to 1, then both $N_0 + 1$ and $N_0 + 2$ may carry appreciable probability.

If $\rho \to \rho_0$ as $L \to \infty$, then the distribution of $\rho(D^* - 2n_0)$ becomes spread out over $\mathbf{Z}$, converging to a non–trivial limit as $L \to \infty$ along any subsequence such that $\phi_0 = \phi_0(L, \rho)$ converges. Both this behaviour and that for larger ρ are quite different from the behaviour found for the [NMW] model in [7]. In the next section, we show that, if ρ becomes smaller, then the differences become less.

3.4 Distance to the continuous circle model

We now show that, as $\rho \to 0$, the distribution of $\rho(D^* - 2n_0)$ approaches the limit distribution T obtained in [7]. Indeed, Theorem 3.10 shows that $\mathbf{P}\left[\rho(D^* - 2n_0) > z\right]$ is close to $\mathbf{E}\{\exp(-e^{2z} W_\rho W'_\rho)\}$. If W_ρ and W'_ρ were replaced by independent standard exponential random variables W and W', the result would be just $\mathbf{P}[T > z]$. Hence our argument is based on showing that the distribution of W_ρ is close to NE (1). We do so by way of Laplace transforms, showing that the Laplace transform

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

of $\mathcal{L}(W_\rho)$ is close to

$$\varphi_e(\theta) \ := \ (1+\theta)^{-1}, \tag{3.22}$$

the Laplace transform of the NE (1) distribution, when $\rho \approx 0$. For this, we employ the characterizing Poincaré equation for Galton–Watson branching processes (see Harris [10], Theorem 8.2, p.15);

$$\varphi_\rho((1+2\rho)\theta) \ = \ f(\varphi_\rho(\theta)). \tag{3.23}$$

We show this to be the fixed point equation corresponding to a contraction Ψ, which also almost fixes φ_e, thus entailing the closeness of φ_ρ and φ_e. The main distributional approximation theorem that results is Theorem 3.16.

To define the operator Ψ, we first need the appropriate function spaces. Let

$$\mathcal{G} \ = \ \left\{ g\colon [0,\infty) \to [0,1] : \ \|g\|_{\mathcal{G}} := \sup_{\theta > 0} \theta^{-2} |g(\theta)| < \infty \right\},$$

and then set

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

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, if χ is the Laplace transform of a random variable X, and if $X_1, X_2, \ldots$ are i.i.d. copies of X, then $\Psi\chi$ is the Laplace transform of $Z = m^{-1}\sum_{i=1}^{N} X_i$, with N, independent of $X_1, X_2, \ldots$, having the distribution $1 + \mathrm{Po}\,(2\rho)$. The Laplace transform φ_ρ of interest to us is a fixed point of Ψ.

Lemma 3.12 *The operator Ψ is a contraction, and, for all $\chi, \psi \in \mathcal{H}$,*

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

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

$$\begin{aligned}
\theta^{-2}|\Psi\chi(\theta) - \Psi\psi(\theta)| \;&=\; \theta^{-2}\left| f\left(\chi\left(\frac{\theta}{m}\right)\right) - f\left(\psi\left(\frac{\theta}{m}\right)\right)\right| \\
&\leq\; \sup_{0\leq t\leq 1} |f'(t)|\;\theta^{-2}\left|\chi\left(\frac{\theta}{m}\right) - \psi\left(\frac{\theta}{m}\right)\right| \\
&=\; \theta^{-2}m\left|\chi\left(\frac{\theta}{m}\right) - \psi\left(\frac{\theta}{m}\right)\right| \\
&=\; m^{-1}(\theta/m)^{-2}\left|\chi\left(\frac{\theta}{m}\right) - \psi\left(\frac{\theta}{m}\right)\right| \\
&\leq\; m^{-1}\|\chi - \psi\|_{\mathcal{G}},
\end{aligned}$$

as required. □

Lemma 3.13 *For the Laplace transform φ_e, we have*

$$\|\Psi\varphi_e - \varphi_e\|_{\mathcal{G}} \;\leq\; \frac{2\rho^2}{(1+2\rho)^2}.$$

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

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

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 $1 + \theta > 1$. □

Lemmas 3.12 and 3.13 together yield the following result.

Lemma 3.14 *For any $\rho > 0$,*

$$\|\varphi_\rho - \varphi_e\|_{\mathcal{G}} \;\le\; \frac{\rho}{1+2\rho}.$$

PROOF: Note that indeed $\varphi_\rho - \varphi_e \in \mathcal{G}$. With Lemmas 3.12 and 3.13, it follows that

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

Thus, since $m > 1$, we obtain

$$\|\varphi_\rho - \varphi_e\|_{\mathcal{G}} \;\le\; \frac{m}{m-1}\frac{2\rho^2}{(1+2\rho)^2} \;=\; \frac{\rho}{1+2\rho},$$

as required. □

As an immediate consequence, $\mathcal{L}(W_\rho) \to \mathrm{NE}(1)$ as $\rho \to 0$. This is the basis of our argument for showing that the distribution of $\rho(D^* - 2n_0)$ is like that of T. What we actually need to compare are the expectations $\mathbf{E}e^{-\theta W_\rho W'_\rho}$ and $\mathbf{E}e^{-\theta W W'}$, for $\theta = e^{2z}$ and any $z \in \mathbf{R}$. The next lemma does this.

Lemma 3.15 *Let W, W' be independent* NE (1) *random variables. Then, for all $\theta > 0$, we have*

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

PROOF: We have

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

Since

$$\mathbf{E}\varphi_e(\theta W'_\rho) \;=\; \mathbf{E}e^{-\theta W W'_\rho} \;=\; \mathbf{E}\varphi_\rho(\theta W),$$

we obtain from the triangle inequality, (3.9) and Lemmas 3.12, 3.13 and 3.14 that

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

as required. □

The quantity $\mathbf{E}e^{-\theta WW'}$ can be more neatly expressed:

$$\mathbf{E}e^{-\theta WW'} = \int_0^\infty \frac{e^{-y}}{1+\theta y}\,dy. \tag{3.24}$$

From this, we obtain the following theorem.

Theorem 3.16 *Let T denote a random variable on $\mathbf{R}$ with distribution given by*

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

Then, for D^* as in (2.7), we have

$$\sup_{z\in\mathbf{R}} |\mathbf{P}[\rho(D^* - 2n_0) > z] - \mathbf{P}[T > z]| = O\left(\rho^{1/3}(1+\log(1/\rho))\right).$$

Proof: We use an argument similar to that used for Theorem 3.10. First, writing $\Delta := D^* - 2n_0$, it follows from (2.7), Lemma 3.15 and (3.24) that

$$\begin{aligned}\left|\mathbf{P}[\rho\Delta > z] - \int_0^\infty \frac{e^{-y}}{1+y\phi_0^2(1+2\rho)^{z/\rho}}\,dy\right| &\le \frac{4\rho}{1+2\rho}(1+2\rho)^{2z/\rho} \\ &\le \frac{4\rho}{1+2\rho}e^{4z}, \qquad z\in\rho\mathbf{Z}.\end{aligned} \tag{3.25}$$

Define $c(\rho)$ by requiring that $(1+2\rho)^{1/\rho} = e^{2c(\rho)}$; then, in view of the fact that $(1+2\rho)^{-1} \le \phi_0 \le 1$, and because, for $a, b > 0$,

$$\left|\int_0^\infty \frac{e^{-y}}{1+ay}\,dy - \int_0^\infty \frac{e^{-y}}{1+by}\,dy\right| \le \frac{|b-a|}{\max\{1,a,b\}}, \tag{3.26}$$

it also follows that

$$\begin{aligned}&\left|\int_0^\infty \frac{e^{-y}}{1+y\phi_0^2(1+2\rho)^{z/\rho}}\,dy - \mathbf{P}[T > zc(\rho)]\right| \\ &\quad\le |\phi_0^2 - 1|(1+2\rho)^{z/\rho} \le \frac{4\rho}{1+2\rho}e^{2z}.\end{aligned} \tag{3.27}$$

Finally, again from (3.26), we have

$$|\mathbf{P}[T > zc(\rho)] - \mathbf{P}[T > z]| \ \leq \ 2|z|(1 - c(\rho)) \min\{1, e^{2zc(\rho)}\}. \tag{3.28}$$

Combining the bounds (3.25), (3.27) and (3.28) for $e^{2z} \leq \rho^{-1/3}$ gives a supremum of order $\rho^{1/3}$ for $|\mathbf{P}[\rho(D^* - 2n_0) > z] - \mathbf{P}[T > z]|$; note that z may actually be allowed to take any real value in this range, since T has bounded density.

For larger values of z, we can use the bound

$$\begin{aligned} \mathbf{P}[T > z] &= \int_0^\infty \frac{e^{-y}dy}{1 + ye^{2z}} \ \leq \ \int_0^1 \frac{dy}{1 + ye^{2z}} \\ &= e^{-2z} \log(1 + e^{2z}) \ \leq \ \rho^{1/3} \log(1 + \rho^{-1/3}), \end{aligned} \tag{3.29}$$

implying 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, as required for Theorem 3.10, we would use $\mathbf{E}\left\{e^{-\phi_0^2(1+2\rho)^{r+r'}W(\tau_r)W'(\tau_r)}\right\}$ in place of $\mathbf{E}\left\{e^{-\phi_0^2(1+2\rho)^{r+r'}W_\rho W'_\rho}\right\}$ to approximate $\mathbf{P}[D > 2n_0 + r + r']$. The distributions of $W(\tau_r)$ and $W'(\tau_r)$ can be calculated iteratively, using the generating function from Lemma 3.3. As D is centred near $2n_0 = 2\lfloor \frac{N}{2} \rfloor$, 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. Shortcuts are realized by taking the union of the ring lattice with a Bernoulli random graph $G_{\Lambda,\frac{\sigma}{\Lambda}}$ having edge probability σ/Λ. 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 comparison with the previous model, in which the k-neighbourhoods on a ring of Lk vertices are replaced by unit intervals on a circle of circumference L, we would define $\rho = k\sigma$, so that the expected number of shortcuts, the edges in $G_{\Lambda,\frac{\sigma}{\Lambda}}$ that are not already in the ring lattice, is close to the value $L\rho/2$ in the previous model. The notation used by [13] is somewhat different.

The model can also be realized by a dynamic construction. Choosing a point $P \in \{1, \ldots, \Lambda\}$ at random, set $R(0) = \{P_0\}$. Then, at the first step (distance 1), the interval consisting of P is increased by k points at each end, and, in addition, a binomially distributed number $M_1^{(1)} \sim \mathrm{Bi}\,(\Lambda - 2k - 1, \frac{\sigma}{\Lambda})$ of shortcuts connect P to centres of new intervals. At each subsequent step, starting from the set $R(n)$ of vertices within distance n of P, each interval is increased by the addition of k points at either end, but with overlapping intervals merged, to form a set $R'(n+1)$. This is then increased to $R(n+1)$

by choosing as shortcuts the edges formed by a random subset of pairs of points, one in $R(n)\setminus R(n-1)$ and one $C\setminus R(n-1)$, each of the possible pairs being independently chosen with probability σ/Λ; those with second end-point in $C \setminus R'(n+1)$ actually contribute new points to $R(n+1)$.

The main complication, compared to the continuous circle model in discrete time, arises from shortcuts having length 1, instead of 0. An interval, when first 'created', consists of a single point. At the next time step, it can start *new* intervals only from shortcuts originating at that single point, whereas, at each subsequent step, there may be as many as $2k$ new points acting as origins for new intervals. Thus there is a one-step hesitation in growth, which makes itself felt if ρ is large. In the same way, because one-point intervals behave differently from all others, the growth and branching analogue of the process needs to have two distinct types of individuals, with type 1 individuals representing one-point intervals, and type 2 individuals representing all others. The quantity corresponding to $(1 + 2\rho)$ now becomes the largest eigenvalue λ of the mean matrix for this two-type branching process.

In this growth and branching analogue, a type 1 interval at time n becomes a type 2 interval at time $n + 1$, increasing its length by k vertices at each end, and, in addition, has a $\mathrm{Bi}\,(\Lambda, \frac{\sigma}{\Lambda})$–distributed number of type 1 intervals as 'offspring'. A type 2 interval at time n stays a type 2 interval at time $n+1$, increasing its length by k vertices at each end, and each of the $2k$ vertices of a type 2 interval that were added at time n is the 'parent' of an independent $\mathrm{Bi}\,(\Lambda, \frac{\sigma}{\Lambda})$–distributed number of type 1 intervals as offspring. Each new interval starts at an independent and uniformly chosen point of the circle, and pairs of parent vertices and their offspring correspond to shortcuts. The initial condition could be a single, randomly chosen point (type 1 interval) P, as above, or a more complicated choice, as in Section 2; a pair of points P and P' at time zero, or just P at time 0, but with a second point P' added at time 1.

In this model, we couple a pair of S- and R-processes by first realizing S, and then realizing R as a sub-process of S, with the help of some extra randomization. In the process R, at each time $n+1$, $n \geq 0$, all shortcuts $\{l, l'\}$ joining $S(n)\setminus S(n-1)$ to $S(n-1)$ are rejected; and if l and l' both belong to $S(n) \setminus S(n-1)$, the shortcut $\{l, l'\}$ is accepted with probability 1/2 if just one of the events $E(l, l'; n)$ and $E(l', l; n)$ has occurred, where

$$E(l, l'; n) \;=\; \{S(n) \setminus S(n-1) \ni l, l';\; l' \text{ is an offspring of } l \text{ at time } n+1\}, \qquad (4.1)$$

and with probability 1 if both of these events have occurred. All descendants of rejected offspring are also rejected, as are the descendants of shortcuts $\{l, l'\}$ joining $S(n)\setminus S(n-1)$ to $S(n+1)$, to avoid a vertex in R having shortcuts sampled more than once. Likewise, when intervals overlap, so that the same vertex in S has offspring assigned more than once, because it is in more than one interval of S, only one set (chosen at random) is accepted for R.

For the growth and branching process S starting from a single point P, write

$$\hat{M}(n) \;:=\; \begin{pmatrix} \hat{M}^{(1)}(n) \\ \hat{M}^{(2)}(n) \end{pmatrix}, \quad n \geq 0,$$

for the numbers of intervals of the two types at time n. Their development over time is given by the branching recursion

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

The total number of intervals at time n is denoted by

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

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

$$(t-1)(t-\sigma) \ = \ 2k\sigma \tag{4.4}$$

of A yields the eigenvalues

$$\lambda \ = \ \lambda_1 \ = \ \tfrac{1}{2}\{\sigma + 1 + \sqrt{(\sigma+1)^2 + 4\sigma(2k-1)}\} \ > \ \sigma + 1; \tag{4.5}$$

$$-\lambda \ < \ \lambda_2 \ = \ \tfrac{1}{2}\{\sigma + 1 - \sqrt{(\sigma+1)^2 + 4\sigma(2k-1)}\} \ < \ 0. \tag{4.6}$$

We note a number of relations involving these eigenvalues, which are useful in what follows. First, from (4.5) and (4.6), we have

$$\lambda_2 \ = \ \sigma + 1 - \lambda, \tag{4.7}$$

and, from (4.4) and (4.5),

$$\lambda - 1 \ \leq \ 2k\sigma; \tag{4.8}$$

then, from (4.5), and since $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ in $a, b \geq 0$, we have

$$\lambda \ \leq \ 1 + \sigma + \sqrt{\sigma(2k-1)}; \tag{4.9}$$

finally, again from (4.4),

$$0 \ \leq \ \omega^2 \ := \ \frac{2k\sigma}{\lambda(\lambda-1)} \ = \ 1 - \sigma/\lambda \ \leq \ 1. \tag{4.10}$$

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

$$f_2^{(i)} \ = \ (\lambda_i - \sigma) f_1^{(i)}. \tag{4.11}$$

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

$$f_1^{(1)} \;=\; (\lambda-\sigma)^{-\frac{1}{2}}, \quad f_2^{(1)} \;=\; (\lambda-\sigma)^{\frac{1}{2}}; \tag{4.12}$$

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 $\sigma(\hat{M}(0),\ldots,\hat{M}(n))$ and the superscript T the vector transpose. Thus, from (4.11),

$$\begin{aligned} W^{(i)}(n) \;&:=\; \lambda_i^{-n}(f^{(i)})^T\hat{M}(n) \\ &=\; \lambda_i^{-n}f_1^{(i)}(\hat{M}_n^{(1)}+(\lambda_i-\sigma)\hat{M}_n^{(2)}) \end{aligned} \tag{4.13}$$

is a (non-zero mean) martingale, for $i=1,2$. Note that

$$\mathbf{E}W^{(i)}(n) \;=\; W_0^{(i)} \;=\; (f^{(i)})^T\hat{M}_0^+ \;=\; (f^{(i)})^T(1,0)^T \;=\; f_1^{(i)} \tag{4.14}$$

for all n, by the martingale property. From (4.13) and (4.11), we also 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

$$\begin{aligned} \hat{M}^{(1)}(n) \;&=\; \lambda^n W^{(1)}(n)\frac{\sigma-\lambda_2}{(\lambda-\lambda_2)f_1^{(1)}}+\lambda_2^n W^{(2)}(n)\frac{\lambda-\sigma}{(\lambda-\lambda_2)f_1^{(2)}}; \\ \hat{M}^{(2)}(n) \;&=\; \lambda^n W^{(1)}(n)\frac{1}{(\lambda-\lambda_2)f_1^{(1)}}-\lambda_2^n W^{(2)}(n)\frac{1}{(\lambda-\lambda_2)f_1^{(2)}}. \end{aligned} \tag{4.15}$$

Define

$$W_{k,\sigma} \;:=\; \lim_{n\to\infty} W^{(1)}(n) \text{ a.s.} \;=\; \lim_{n\to\infty}\lambda_1^{-n}(f^{(1)})^T\hat{M}(n) \text{ a.s.} \tag{4.16}$$

to 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 5.9 and 5.12. Let D_d^* denote a random variable on the integers with distribution given by

$$\mathbf{P}\left[D_d^* > 2n_d+x\right] \;=\; \mathbf{E}\exp\left\{-\frac{\lambda^2}{(\lambda-\lambda_2)}(\lambda-\sigma)\phi_d^2\lambda^x W_{k,\sigma}W'_{k,\sigma}\right\}, \tag{4.17}$$

for any $x\in\mathbf{Z}$. Here, n_d and ϕ_d are such that $\lambda^{n_d}=\phi_d(\Lambda\sigma)^{1/2}$ and $\lambda^{-1}<\phi_d\le 1$, and $W'_{k,\sigma}$ is an independent copy of $W_{k,\sigma}$. Let D_d denote the graph distance between a randomly chosen pair of vertices P and P' on the ring lattice C.

Theorem 4.1 *If $\Lambda\sigma \to \infty$ and $\rho = k\sigma$ remains bounded, then $d_K(\mathcal{L}(D_d), \mathcal{L}(D_d^*)) \to 0$. If $\rho \to 0$, then $\rho(D_d^* - 2n_d) \to_{\mathcal{D}} T$, where T is as in Theorem 2.1.*

Note that, in (4.17), the expectation is taken with respect to independent random variables $W_{k,\sigma}$ and $W'_{k,\sigma}$, each having the distribution of the martingale limit $\lim_{n\to\infty} W^{(1)}(n)$ conditional on the initial condition $\hat{M}(0) = \mathbf{e}^{(1)}$. We shall later need also to consider the distribution of $W_{k,\sigma}$ under other initial conditions.

5 The discrete circle model: proofs

As in the previous section, we run the growth and branching process starting from two initial points P and P', the second either being present at time 0 or else only at time 1; from the branching property, we can regard the process as the sum of two independent processes $\hat{M}$ and $\hat{N}$ having the same distribution, with $\hat{N}$ started either at time 0 or at time 1. We then investigate, at times $m \geq 1$, whether or not there are intervals of $\hat{M}$ and $\hat{N}$ that overlap in a way which implies that there are points of R which can be reached from P within distance m and from P' within distance m or $m-1$, implying that the distance D between P and P' is no greater than $2m$ or $2m-1$, respectively. It is convenient to write the times m in the form $n_d + r$, where n_d is such that $\lambda^{n_d} \leq (\Lambda\sigma)^{1/2} < \lambda^{n_d+1}$, using notation of the form $\hat{M}_r$ to denote $\hat{M}(n_d + r)$, and we write $2\tau_r := \{2k(n_d + r) + 1\}$ for the length of the longest possible interval in the growth and branching process at time $n_d + r$.

5.1 Poisson approximation for the intersection probability

The first step is to find an approximation to the event $A_{r,r'} := \{D_d > 2n_d + r + r'\}$ — where, as before, $r' = r$ if the process $\hat{N}$ starts at time 0, and $r' = r - 1$ if it starts at time 1 — which can be represented as the event that a dissociated sum of indicator random variables takes the value 0, enabling the Stein–Chen method to be used. If, at time $n_d + r$, the $\hat{M}_r^+ := \hat{M}_r^{(1)} + \hat{M}_r^{(2)}$ intervals of $\hat{M}$ and the $\hat{N}_r^+ := \hat{N}_r^{(1)} + \hat{N}_r^{(2)}$ intervals of $\hat{N}$ are $I_1, I_2, \ldots, I_{\hat{M}_r^+}$ and $J_1, J_2, \ldots, J_{\hat{N}_r^+}$ respectively, we define

$$V_{r,r'} \;:=\; \sum_{i=1}^{\hat{M}_r^+} \sum_{j=1}^{\hat{N}_r^+} X_{ij},$$

much as in (2.5), where

$$X_{ij} \;:=\; \mathbf{1}\{I_i \cap J_j \neq \emptyset\}\mathbf{1}\{I_i^{2k} \not\subset J_j\}\mathbf{1}\{J_j^{2k} \not\subset I_i\}(1 - K_{ij})(1 - K'_{ji}), \quad i, j \geq 1, \qquad (5.1)$$

and

$$K_{ij} \;:=\; \mathbf{1}\{I_i^k \subset J_j,\, Z_i = 0\}; \quad K'_{ji} \;:=\; \mathbf{1}\{J_j^k \subset I_i,\, Z'_j = 0\},$$

and where $(Z_i,\, i \geq 2)$ and $(Z'_j,\, j \geq 2)$ are sequences of independent Bernoulli Be $(1/2)$ random variables; for an interval $K = [a, b]$, the notation K^l once again denotes the interval $[a-l, b+l]$. This definition of $V_{r,r'}$ has some slight differences from that previously defined in (2.6), occasioned largely because shortcuts are now taken to have length 1, rather than 0. It does not count overlaps which result from links between $S(n) \setminus S(n-1)$ and $S(n-1)$, $n \geq 1$, which can be distinguished at time $n_d + r$ because they give rise to intervals I_i and J_j such that either $\{I_i^{2k} \subset J_j\}$ or $\{J_j^{2k} \subset I_i\}$; and it only counts links between pairs of points in $S(n) \setminus S(n-1)$ with probability 1/2. These provisions ensure that the event $\{V_{r,r'} = 0\}$ is close enough to $A_{r,r'}$, but the two events are not exactly the same. The definition of X_{ij} does not exclude any descendants of rejected intervals from consideration, so that $V_{r,r'}$ counts more intersections than are actually represented in R. Furthermore, with this definition of X_{ij}, if two vertices l, l' are in $S(n) \setminus S(n-1)$ for some n, and $E(l, l'; n)$ and $E(l', l; n)$ (as defined in (4.1)) both occur, then the pair contributes $0, 1$ or 2 to $V_{r,r'}$, depending on the values of the corresponding indicators Z_i and Z'_j, instead of the value 1 for the single intersection in R. Finally, note that, if $r' = r-1$, the random variable Z'_1 needs to be defined, since it is possible to have $J_1^k \subset I_1$ (when $P' = P$); it is then correct to take $Z'_1 = 1$ a.s., since the point P' is never rejected.

To address the possibility of there being pairs l, l' for which $E(l, l'; n)$ and $E(l', l; n)$ both occur, we define the event

$$E_{r,r'} := \bigcup_{i=1}^{\hat{M}_r^+} \bigcup_{j=1}^{\hat{N}_r^+} \mathbf{1}\{|I_i| = |J_j|\}\, \mathbf{1}\{\zeta_i = \xi'_j\}\, \mathbf{1}\{\zeta'_j = \xi_i\}, \tag{5.2}$$

where ζ_i, ξ_i and ζ'_j, ξ'_j denote the centres and parent vertices of the intervals I_i and J_j, respectively: $E_{r,r'}$ is the event that there is some edge $\{l, l'\}$ chosen simultaneously by both growth and branching processes before time $n_d + r$. Then, writing $\mathcal{F}_{r,r'}$ for the σ-algebra $\sigma\{(\hat{M}(l), \hat{N}(l)),\ 0 \leq l \leq n_d + r\}$, we have

$$\mathbf{P}[E_{r,r'} \,|\, \mathcal{F}_{r,r'}] = \sum_{l=1}^{n_d+r} \hat{M}^{(1)}(l) \hat{N}^{(1)}(l) \Lambda^{-2}, \tag{5.3}$$

because two centres must coincide with specific points of C. The possible overcounting in $V_{r,r'}$ is accounted for just as in the proof of Lemma 3.2 by defining a corresponding sum $\widetilde{V}_{r,r'}$, and then using the relationship

$$\{V_{r,r'} = 0\} \setminus E_{r,r'} \subset A_{r,r'} \subset \{\widetilde{V}_{r,r'} = 0\} \cup E_{r,r'}, \tag{5.4}$$

in place of (3.1). This leads to the following analogue of Lemma 3.2. The proof is very similar, and is therefore omitted.

Lemma 5.1 *With the above assumptions and definitions, it follows that*

$$\begin{aligned} &|\mathbf{P}[A_{r,r'}] - \mathbf{P}[V_{r,r'} = 0]| \\ &\quad \leq 16\tau_r^2 \Lambda^{-2} \mathbf{E}\{\hat{M}_r^+ \hat{N}_r^+ (\hat{M}_r^+ + \hat{N}_r^+)(1 + \log \hat{M}_r^+ + \log \hat{N}_r^+)\} \\ &\qquad\qquad + \Lambda^{-2} \sum_{l=1}^{n_d+r} \mathbf{E}\{\hat{M}^{(1)}(l) \hat{N}^{(1)}(l)\}. \end{aligned}$$

Then, defining

$$\hat{\lambda}_r(\hat{M}, \hat{N}) \;:=\; \sum_{i=1}^{\hat{M}_r^+} \sum_{j=1}^{\hat{N}_r^+} \mathbf{E}\{X_{ij} \,|\, \mathcal{F}_{r,r'}\}, \tag{5.5}$$

an argument similar to that of Proposition 3.1 establishes the following result.

Proposition 5.2 *We have*

$$|\mathbf{P}[V_{r,r'} = 0 \,|\, \mathcal{F}_{r,r'}] - \exp\{-\hat{\lambda}_r(\hat{M}, \hat{N})\}| \;\le\; 8\tau_r \Lambda^{-1}(\hat{M}_r^+ + \hat{N}_r^+). \qquad \Box$$

The next step is to establish fomulae for the moments appearing in the bounds of the previous two results. The corresponding results are much more complicated that those of Lemma 3.3, and their proofs are deferred to the appendix.

Lemma 5.3 *For the means,*

$$\begin{aligned} \mathbf{E}\hat{M}^{(1)}(n) &= \frac{1}{\lambda - \lambda_2}(\lambda^n(\sigma - \lambda_2) + \lambda_2^n(\lambda - \sigma)) \;\le\; \lambda^n; &(5.6)\\ \mathbf{E}\hat{M}^{(2)}(n) &= \frac{1}{\lambda - \lambda_2}(\lambda^n - \lambda_2^n), \end{aligned}$$

and

$$\begin{aligned} \mathbf{E}(\hat{M}^{(1)}(n) + 2k\hat{M}^{(2)}(n)) &= \frac{1}{(\lambda - \lambda_2)}\{(1 - \lambda_2/\sigma)\lambda^{n+1} + (\lambda_2/\sigma)(\lambda - \sigma)\lambda_2^n\} \\ &\le\; 2k\lambda^n. &(5.7) \end{aligned}$$

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

$$\begin{aligned} &\mathrm{Var}\,(W^{(1)}(j) - W^{(1)}(n)) \;\le\; \omega^2(\lambda - \sigma)^{-1}\lambda^{-j}; &(5.8)\\ &\mathrm{Var}\,(W^{(2)}(j) - W^{(2)}(n)) &(5.9)\\ &\quad\le\; 2k\sigma\left(\frac{1}{\lambda_2\sqrt{\sigma - \lambda_2}}\right)^2 \min\left\{\frac{\lambda_2^2}{|\lambda - \lambda_2^2|}, (n-j)\right\}\left(\frac{\lambda}{\lambda_2^2}\right)^j \max\left\{1, \frac{\lambda}{\lambda_2^2}\right\}^{n-j}. \end{aligned}$$

Then, for $\hat{M}^+(n)$, we have

$$\begin{aligned} \mathbf{E}\hat{M}^+(n) &= \frac{1}{\lambda - \lambda_2}(\lambda^{n+1} - \lambda_2^{n+1}) \;\le\; 2\lambda^n; &(5.10)\\ \mathrm{Var}\,\hat{M}^+(n) &\le\; 4\omega^2\lambda^{2n}; \quad \mathbf{E}\{\hat{M}^+(n)\}^2 \;\le\; 4(1+\omega^2)\lambda^{2n}; &(5.11)\\ \mathbf{E}\{\hat{M}^+(n)\log\hat{M}^+(n)\} &\le\; 2(4 + n\log\lambda)\lambda^n; &(5.12) \end{aligned}$$

and

$$\mathbf{E}\{(\hat{M}^+(n))^2\log\hat{M}^+(n)\} \;\le\; 8(5 + n\log\lambda)\lambda^{2n}. \tag{5.13}$$

Applying the bounds in the preceding lemma to Lemma 5.1 and Proposition 5.2, we deduce the following approximation.

Lemma 5.4 *With the above notation and definitions, we have*

$$\begin{aligned} &|\mathbf{P}[D_d > 2n_d + r + r'] - \mathbf{E}(\exp\{-\hat{\lambda}_r(\hat{M}, \hat{N})\})| \\ &\quad\le\; 256\tau_r^2\Lambda^{-2}\lambda^{3n_d+2r+r'}\{14 + 3(n_d + r)\log\lambda\} \\ &\qquad + \Lambda^{-2}(\lambda - 1)^{-1}\lambda^{2n_d+r+r'+1} + 32\tau_r\Lambda^{-1}\lambda^{n_d+r}. \end{aligned}$$

5.2 The shortest path length: error bounds

In order to make use of Lemma 5.4, we now need a simpler expression in place of $\hat\lambda_r(\hat M,\hat N)$. We begin by deriving a computable representation in terms of the processes $\hat M$ and $\hat N$.

Lemma 5.5 *We have the expression*

$$\begin{aligned}\Lambda\hat\lambda_r(\hat M,\hat N) = 4k &\sum_{l=0}^{n_d+r-1} \hat M^+(l)\hat N^+(l) + 3k\hat M_r^+\hat N_r^+\\ &- (k-\tfrac12)\sum_{l=1}^{n_d+r}\Big\{\hat M^{(1)}(l)\hat N^{(1)}(l-1)+\hat M^{(1)}(l-1)\hat N^{(1)}(l)\Big\}\\ &- (3k-1)\sum_{l=0}^{n_d+r}\hat M^{(1)}(l)\hat N^{(1)}(l) + \tfrac12\mathbf{1}\{\hat N^{(1)}(0)=0,\ \hat N^{(1)}(1)=1\}.\end{aligned}$$

Proof: In view of the definition (5.5) of $\hat\lambda_r(\hat M,\hat N)$, we first need an expression for $\mathbf{E}\{X_{ij}\,|\,\mathcal{F}_{r,r'}\}$. This is a function of the lengths s_i and u_j of the corresponding intervals I_i and J_j, quantities which are indeed $\mathcal{F}_{r,r'}$–measurable. Writing $m_{ij} := \min(s_i,u_j)$, and recalling that both s_i and u_j belong to the set $2k\mathbf{Z}+1$, it follows from the definition (5.1) of X_{ij} that

$$\mathbf{E}\{X_{ij}\,|\,|I_i|=s_i,|J_j|=u_j\} \;=\; \Lambda^{-1}\begin{cases}2(m_{ij}-1)+3k & \text{if } |s_i-u_j|\ge 4k;\\ 2(m_{ij}-1)+2k+\frac12 & \text{if } |s_i-u_j|=2k;\\ 2(m_{ij}-1)+1 & \text{if } s_i=u_j,\end{cases}$$

except that, if $r'=r-1$, then X_{11} is special:

$$\mathbf{E}\{X_{11}\,|\,\mathcal{F}_{r,r'}\} \;=\; \Lambda^{-1}(2(m_{11}-1)+2k+1).$$

Hence, rewriting the sum defining $\hat\lambda_r(\hat M,\hat N)$ in terms of the steps of the processes $\hat M$ and $\hat N$, we find that

$$\begin{aligned}\Lambda\hat\lambda_r(\hat M,\hat N) \;=\; &\sum_{s=0}^{n_d+r}\sum_{s'=0}^{n_d+r}\hat M^{(1)}(s)\hat N^{(1)}(s')\\ &\times\big\{2(2\tau_r-1-2k\max(s,s'))+3k-(k-\tfrac12)\mathbf{1}\{|s-s'|=1\}-(3k-1)\mathbf{1}\{s=s'\}\big\}\\ &\qquad+\tfrac12(r-r')\\ =\; &\sum_{s=2}^{n_d+r}\Big(\hat M^{(1)}(s)\hat N^+(s-2)+\hat M^+(s-2)\hat N^{(1)}(s)\Big)\{2(2\tau_r-1-2ks)+3k\}\\ &+\sum_{s=1}^{n_d+r}\Big(\hat M^{(1)}(s)\hat N^{(1)}(s-1)+\hat M^{(1)}(s-1)\hat N^{(1)}(s)\Big)\{2(2\tau_r-1-2ks)+2k+\tfrac12\}\\ &+\sum_{s=0}^{n_d+r}\hat M^{(1)}(s)\hat N^{(1)}(s)\{2(2\tau_r-1-2ks)+1\}+\tfrac12(r-r')\end{aligned}$$

$$
\begin{aligned}
&= \sum_{s=1}^{n_d+r} \nabla\{\hat{M}^+(s)\hat{N}^+(s)\}\{2(2\tau_r - 1 - 2ks) + 3k\} \\
&\quad - (k - \tfrac{1}{2}) \sum_{s=1}^{n_d+r} \Big(\hat{M}^{(1)}(s)\hat{N}^{(1)}(s-1) + \hat{M}^{(1)}(s-1)\hat{N}^{(1)}(s)\Big) \\
&\quad - (3k-1) \sum_{s=0}^{n_d+r} \hat{M}^{(1)}(s)\hat{N}^{(1)}(s) + \tfrac{1}{2}(r - r'),
\end{aligned}
$$

where we also use the fact that $\hat{M}^+(l-1) = \hat{M}^{(2)}(l)$ for each l. The lemma follows by partial summation. $\square$

Armed with this expression for $\hat{\lambda}_r(\hat{M},\hat{N})$, we can now derive a more tractable approximation. We introduce the notation

$$
\gamma \;:=\; \gamma(k,\sigma) \;:=\; \min\{\tfrac{1}{2}, (\log(\lambda/|\lambda_2|)/\log\lambda\}. \tag{5.14}
$$

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 $\gamma = \frac{1}{2}$.

Lemma 5.6 *We have the approximation*

$$
\begin{aligned}
&\mathbf{E}\left|\hat{\lambda}_r(\hat{M},\hat{N}) \;-\; \frac{\lambda^2}{(\lambda-\lambda_2)}(\lambda-\sigma)(\Lambda\sigma)^{-1}\lambda^{2n_d+r+r'}W_{k,\sigma}W'_{k,\sigma}\right| \\
&\quad\leq\; (\Lambda\sigma)^{-1}\left\{11\lambda(\lambda-\sigma)\lambda^{(2-\gamma)n_d+r+r'-\gamma r'}\{1+\sqrt{(\lambda-1)(n_d+r+1)}\} + \lambda^2\right\}.
\end{aligned}
$$

Proof: We give the argument only for the case $r' = r-1$, which is slightly more complicated; note that, in this situation, $\mathcal{L}(\hat{N}(s)) = \mathcal{L}(\hat{M}(s-1))$ for $s \geq 1$, and that $\hat{N}(0) = 0$. We begin by observing that, from (4.15), (5.8), (4.7) and (5.9), and from the definition of ω in (4.10), we have

$$
\begin{aligned}
&\mathbf{E}\left|\hat{M}^{(1)}(s) \;-\; \frac{\lambda^s W_{k,\sigma}(\sigma-\lambda_2)}{(\lambda-\lambda_2)f_1^{(1)}}\right| \\
&= \;\mathbf{E}\left|\frac{\lambda^s(W^{(1)}(s)-W_{k,\sigma})(\sigma-\lambda_2)}{(\lambda-\lambda_2)f_1^{(1)}} + \frac{\lambda_2^s(W^{(2)}(s)-f_1^{(2)})(\lambda-\sigma)}{(\lambda-\lambda_2)f_1^{(2)}} + \frac{\lambda_2^s(\lambda-\sigma)}{\lambda-\lambda_2}\right| \\
&\leq\; \frac{\lambda^s(\lambda-1)}{\lambda-\lambda_2}\,\omega\lambda^{-s/2} \\
&\qquad + \frac{|\lambda_2|^s(\lambda-\sigma)}{\lambda-\lambda_2}\left(1+\sqrt{2k\sigma\min\left\{\frac{s}{\lambda_2^2},\frac{1}{|\lambda-\lambda_2^2|}\right\}}\right)\max\left\{1,\frac{\lambda}{|\lambda_2|^2}\right\}^{s/2}.
\end{aligned} \tag{5.15}
$$

Now, because $\min\{x^{-1}s, |\lambda-x|^{-1}\} \le \lambda^{-1}(s+1)$ for $x > 0$ and because, from the definition of γ in (5.14), we have $\max\{\lambda^{1/2}, |\lambda_2|\} \le \lambda^{1-\gamma}$, it follows from (5.15) that

$$\mathbf{E}\left|\hat{M}^{(1)}(s) \;-\; \frac{\lambda^s W_{k,\sigma}(\sigma-\lambda_2)}{(\lambda-\lambda_2)f_1^{(1)}}\right| \;\le\; \frac{\lambda^{s(1-\gamma)}}{\lambda-\lambda_2}\left\{(\lambda-1)+(\lambda-\sigma)\left(1+\sqrt{\frac{2k\sigma(s+1)}{\lambda}}\right)\right\};$$

using (4.10) and (4.7), together with the fact that $\lambda_2 < 0 < \sigma$, it now follows that

$$\mathbf{E}\left|\hat{M}^{(1)}(s) \;-\; \frac{\lambda^s W_{k,\sigma}(\sigma-\lambda_2)}{(\lambda-\lambda_2)f_1^{(1)}}\right| \;\le\; \lambda^{s(1-\gamma)}\{1+\sqrt{(1+s)(\lambda-1)}\} \;=:\; \chi(s). \tag{5.16}$$

By a similar argument, we also obtain

$$\begin{aligned}
&\mathbf{E}\left|\hat{M}^{+}(s) \;-\; \frac{\lambda^{s+1} W_{k,\sigma}}{(\lambda-\lambda_2)f_1^{(1)}}\right| \\
&\;=\; \mathbf{E}\left|\frac{\lambda^{s+1}(W^{(1)}(s)-W_{k,\sigma})}{(\lambda-\lambda_2)f_1^{(1)}} - \frac{\lambda_2^{s+1}(W^{(2)}(s)-f_1^{(2)})}{(\lambda-\lambda_2)f_1^{(2)}} - \frac{\lambda_2^{s+1}}{\lambda-\lambda_2}\right| \\
&\;\le\; \frac{\lambda^{s+1}}{\lambda-\lambda_2}\omega\lambda^{-s/2} + \frac{|\lambda_2|^{s+1}}{\lambda-\lambda_2}\left(1+\sqrt{2k\sigma\min\left\{\frac{s}{\lambda_2^2}, \frac{1}{|\lambda-\lambda_2^2|}\right\}}\right)\max\left\{1, \frac{\lambda}{|\lambda_2|^2}\right\}^{s/2} \\
&\;\le\; \frac{\lambda^{s(1-\gamma)}}{\lambda-\lambda_2}\left\{\lambda\omega + |\lambda_2|\left(1+\sqrt{\frac{2k\sigma(s+1)}{\lambda}}\right)\right\} \\
&\;\le\; \lambda^{s(1-\gamma)}\{1+\sqrt{(1+s)(\lambda-1)}\} \;=\; \chi(s);
\end{aligned} \tag{5.17}$$

the corresponding results for $\hat{N}(s)$ follow from these, since $\mathcal{L}(\hat{N}(s)) = \mathcal{L}(\hat{M}(s-1))$ for $s \ge 1$. Hence, and from (5.6) and (5.10), noting also that

$$\mathbf{E}\left\{\frac{\lambda^s W_{k,\sigma}(\sigma-\lambda_2)}{(\lambda-\lambda_2)f_1^{(1)}}\right\} \;=\; \frac{\lambda^s(\lambda-1)}{\lambda-\lambda_2} \;\le\; \lambda^s$$

and that

$$\mathbf{E}\left\{\frac{\lambda^{s+1} W_{k,\sigma}}{(\lambda-\lambda_2)f_1^{(1)}}\right\} \;=\; \frac{\lambda^{s+1}}{\lambda-\lambda_2} \;\le\; \lambda^s,$$

we have, for example, for substitution into Lemma 5.5, the inequalities

$$\begin{aligned}
&\mathbf{E}\left|\hat{M}^{+}(l)\hat{N}^{+}(l) - \frac{\lambda^{2l+1} W_{k,\sigma}W'_{k,\sigma}}{(\lambda-\lambda_2)^2(f_1^{(1)})^2}\right| \\
&\;\le\; \mathbf{E}\{\hat{N}^{+}(l)\}\mathbf{E}\left|\hat{M}^{+}(l) - \frac{\lambda^{l+1}W_{k,\sigma}}{(\lambda-\lambda_2)f_1^{(1)}}\right| + \mathbf{E}\left\{\frac{\lambda^{l+1}W_{k,\sigma}}{(\lambda-\lambda_2)f_1^{(1)}}\right\}\mathbf{E}\left|\hat{N}^{+}(l) - \frac{\lambda^{l}W'_{k,\sigma}}{(\lambda-\lambda_2)f_1^{(1)}}\right| \\
&\;\le\; 2\lambda^{l-1}\chi(l) + \lambda^l\chi(l-1)
\end{aligned}$$

and

$$\mathbf{E}\left|\hat{M}^{(1)}(l)\hat{N}^{(1)}(l-1) - \frac{\lambda^{2l-2}W_{k,\sigma}W'_{k,\sigma}(\sigma-\lambda_2)^2}{(\lambda-\lambda_2)^2(f_1^{(1)})^2}\right| \;\le\; \lambda^{l-2}\chi(l) + \lambda^l\chi(l-2).$$

Thus it follows from Lemma 5.5 that

$$\begin{aligned}
&\mathbf{E}|\hat{\lambda}_r(\hat{M},\hat{N}) - \Lambda^{-1}W_{k,\sigma}W'_{k,\sigma}C_1(k,\sigma,r)| && (5.18)\\
&\le\; 4k\Lambda^{-1}\sum_{s=1}^{n_d+r-1}\{2\lambda^{s-1}\chi(s)+\lambda^s\chi(s-1)\} \\
&\quad + 3k\Lambda^{-1}\{2\lambda^{n_d+r-1}\chi(n_d+r)+\lambda^{n_d+r}\chi(n_d+r-1)\} \\
&\quad + (k-\tfrac{1}{2})\Lambda^{-1}\left\{1+\sum_{s=2}^{n_d+r}\{\lambda^{s-2}\chi(s)+\lambda^s\chi(s-2)+2\lambda^{s-1}\chi(s-1)\}\right\} \\
&\quad + (3k-1)\Lambda^{-1}\sum_{s=1}^{n_d+r}\{\lambda^{s-1}\chi(s)+\lambda^s\chi(s-1)\} + \tfrac{1}{2}\Lambda^{-1}, && (5.19)
\end{aligned}$$

where

$$\begin{aligned}
&\{(\lambda-\lambda_2)f_1^{(1)}\}^2C_1(k,\sigma,r) \\
&:=\; 4k\sum_{s=1}^{n_d+r-1}\lambda^{2s+1} + 3k\lambda^{2(n_d+r)+1} - (2k-1)\sum_{s=2}^{n_d+r}\lambda^{2s-2}(\sigma-\lambda_2)^2 \\
&\qquad - (3k-1)\sum_{s=1}^{n_d+r}\lambda^{2s-1}(\sigma-\lambda_2)^2 \\
&=\; \frac{\lambda^{2(n_d+r)}}{\lambda^2-1}\{4k\lambda + 3k\lambda(\lambda^2-1) - (\sigma-\lambda_2)^2(2k-1+\lambda(3k-1))\} \\
&\qquad - \frac{\lambda}{\lambda^2-1}\{4k\lambda^2 - (\sigma-\lambda_2)^2(\lambda(2k-1)+3k-1)\}.
\end{aligned}$$

The expression for $\{(\lambda-\lambda_2)f_1^{(1)}\}^2C_1(k,\sigma,r)$ simplifies astonishingly. For the coefficient of $\lambda^{2(n_d+r)}$, using (4.4) and (4.7) to express it in terms of λ and σ alone, one obtains $\lambda\sigma^{-1}(2\lambda-\sigma-1)$, which is just $\lambda\sigma^{-1}(\lambda-\lambda_2)$, again by (4.7). By the same strategy, the remaining term yields

$$\lambda(2\sigma)^{-1}\{\lambda(2\lambda^2-7\lambda+3)-\sigma(2\lambda^2-5\lambda+1)\},$$

which, since $0<\sigma<\lambda$, is in modulus less than $\sigma^{-1}\lambda^4$ in $\lambda\ge 1$. This implies that

$$|\{(\lambda-\lambda_2)f_1^{(1)}\}^2C_1(k,\sigma,r) - (\lambda-\lambda_2)\sigma^{-1}\lambda^{2n_d+r+r'+2}| \;\le\; \sigma^{-1}\lambda^4, \tag{5.20}$$

so that

$$\Lambda^{-1}\mathbf{E}W_{k,\sigma}W'_{k,\sigma}\left|C_1(k,\sigma,r) - \frac{\lambda^2(\lambda-\sigma)}{\sigma(\lambda-\lambda_2)}\lambda^{2n_d+r+r'}\right| \;\le\; \Lambda^{-1}(\mathbf{E}W_{k,\sigma})^2\frac{\lambda^4(\lambda-\sigma)}{\sigma(\lambda-\lambda_2)^2} \;\le\; (\Lambda\sigma)^{-1}\lambda^2. \tag{5.21}$$

Next, substitute the expression (5.16) into the right-hand side of (5.19), and simplify. We use the simple bound

$$\chi(s) \;\le\; \lambda^{s(1-\gamma)}\{1+\sqrt{(\lambda-1)(n_d+r+1)}\},$$

express k in terms of λ and σ using (4.4), and note that

$$\frac{\lambda^{(2-\gamma)(n_d+r)}(\lambda-1)}{\lambda^{2-\gamma}-1} \;=\; \lambda^{(2-\gamma)n_d+r+r'-\gamma r'}\cdot\frac{\lambda^{1-\gamma}(\lambda-1)}{\lambda^{2-\gamma}-1} \;\le\; \lambda^{(2-\gamma)n_d+r+r'-\gamma r'},$$

and that

$$\{\tfrac{3}{2}(\lambda-1)+2\}\{2\lambda^{-\gamma}+1\} \;\le\; 6\lambda$$

in $\lambda \ge 1$; collecting terms, we obtain at most

$$11(\Lambda\sigma)^{-1}\lambda(\lambda-\sigma)\lambda^{(2-\gamma)n_d+r+r'-\gamma r'}\{1+\sqrt{(\lambda-1)(n_d+r+1)}\}, \tag{5.22}$$

and the lemma follows from (5.19), (5.21) and (5.22). $\square$

Lemma 5.6 can now be combined with Lemma 5.4 to obtain our main approximation for the shortest path length D_d. Recall that n_d is such that $\lambda^{n_d} = \phi_d(\Lambda\sigma)^{1/2}$, where we require $\lambda^{-1} < \phi_d \le 1$, so that Lemma 5.6 approximates $\hat\lambda_r(\hat M,\hat N)$ by

$$\frac{\lambda^2}{(\lambda-\lambda_2)}(\lambda-\sigma)\phi_d^2\lambda^{r+r'}W_{k,\sigma}W'_{k,\sigma},$$

a quantity appearing in the definition (4.17) of $\mathcal{L}(D_d^*)$. Recall also that $2\tau_r = 2k(n_d+r)+1$, and define the quantities

$$\begin{aligned}
\eta_1'(r,r') &:= 64\phi_d^3\{\sigma(2k(n_d+r)+1)\}^2\lambda^{2r+r'}\{14+3(n_d+r)\log\lambda\};\\
\eta_2'(r,r') &:= 16\phi_d\{\sigma(2k(n_d+r)+1)\}\lambda^{r};\\
\eta_3'(r,r') &:= \tfrac{1}{2}\phi_d^2\sigma\lambda^{r+r'+2}+\lambda^2,
\end{aligned}$$

and

$$\eta_4'(r,r') \;:=\; 11\phi_d^{2-\gamma}\lambda(\lambda-\sigma)\{1+\sqrt{(\lambda-1)(n_d+r+1)}\}\lambda^{r+r'-\gamma r'},$$

which are used to express the approximation. The quantity $\eta_1'(r,r')$ comes from the approximation of $A_{r,r'}$ by $\{V_{r,r'}=0\}$ in Lemma 5.1, $\eta_2'(r,r')$ comes from the Poisson approximation of Proposition 5.2, and $\eta_4'(r,r')$ comes from the approximation of $\hat\lambda_r(\hat M,\hat N)$ in Lemma 5.6; $\eta_3'(r,r')$ relates to elements of smaller order arising in Lemmas 5.1 and 5.6. Then we have the following analogue of Theorem 3.8.

Theorem 5.7 *With the above assumptions and definitions, for $x \in \mathbf{Z}$ and $r' = r'(x) = \lfloor x/2 \rfloor$, $r = r(x) = x - r'(x) \le (x+1)/2$, we have*

$$\begin{aligned}
&|\mathbf{P}[D_d > 2n_d + x] - \mathbf{P}[D_d^* > 2n_d + x]|\\
&\quad\le\; \{\eta_1'(r,r')+\eta_2'(r,r')\}(\Lambda\sigma)^{-1/2}+\eta_3'(r,r')(\Lambda\sigma)^{-1}+\eta_4'(r,r')(\Lambda\sigma)^{-\gamma/2}.
\end{aligned}$$

In particular, if $\rho = k\sigma \le 1$, then

$$\begin{aligned}
&|\mathbf{P}[D_d > 2n_d + x] - \mathbf{P}[D_d^* > 2n_d + x]|\\
&\quad\le\; \{\eta_1'(r,r')+\eta_2'(r,r')\}(\Lambda\sigma)^{-1/2}+\eta_3'(r,r')(\Lambda\sigma)^{-1}+\eta_4'(r,r')(\Lambda\sigma)^{-1/4}.
\end{aligned}$$

The theorem can be translated into a uniform bound, similar to that of Theorem 3.10. 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 3.9 makes this possible. To state it, we first need some notation.

For $W_{k,\sigma}$ as in (4.16), let $\varphi = \varphi_{k,\sigma} = (\varphi^{(1)}, \varphi^{(2)})$ denote the Laplace transforms

$$\begin{aligned} \varphi^{(1)}(\theta) &= \mathbf{E}\left\{e^{-\theta\sqrt{\lambda-\sigma}\,W_{k,\sigma}} \mid \hat{M}(0) = \mathbf{e}^{(1)}\right\}; \qquad (5.23) \\ \varphi^{(2)}(\theta) &= \mathbf{E}\left\{e^{-\theta\sqrt{\lambda-\sigma}\,W_{k,\sigma}} \mid \hat{M}(0) = \mathbf{e}^{(2)}\right\} \end{aligned}$$

of $\mathcal{L}(\sqrt{\lambda-\sigma}\,W_{k,\sigma})$, 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, φ satisfies the Poincaré equation

$$\varphi^{(i)}(\lambda\theta) = g^i(\varphi^{(1)}(\theta), \varphi^{(2)}(\theta)) \quad \text{in} \quad \Re\theta \geq 0; \quad i = 1, 2, \qquad (5.24)$$

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,

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

and

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

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

Lemma 5.8 *For $\theta, \sigma > 0$, we have*

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

and hence

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

Proof: Put

$$\varphi^{(i)}_n(\theta) = \mathbf{E}\left(e^{-\theta\sqrt{\lambda-\sigma}\,W^{(1)}(n)} \mid \hat{M}(0) = \mathbf{e}^{(i)}\right), \quad i = 1, 2.$$

We proceed by induction on n. First, we have

$$\begin{aligned} \varphi_0^{(1)}(\theta) &= e^{-\theta} \leq \frac{1}{1+\theta}; \\ \varphi_0^{(2)}(\theta) &= e^{-\theta(\lambda-\sigma)} \leq \frac{1}{1+\theta(\lambda-\sigma)}. \end{aligned}$$

Assume that

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

By the Poincaré recursion,

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

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

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

and, also from (4.4),

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

Taking limits as $n \to \infty$ proves the first two assertions. The last assertion follows from Lemma 3.9. $\square$

Theorem 5.9 *For D_d the shortest path length and D_d^* with distribution given by (4.17), we have*

$$d_K(\mathcal{L}(D_d), \mathcal{L}(D_d^*)) \;=\; O\left((\Lambda\sigma)^{-\gamma/(4-\gamma)} \log(\Lambda\sigma)\right),$$

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 (5.14). Hence $d_K(\mathcal{L}(D_d), \mathcal{L}(D_d^*)) \to 0$ 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 $0 < G < 1$, 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 we have

$$(n_d + r(x))\log\lambda \;=\; O\left(\log(\Lambda\sigma)\right), \qquad \sigma\{2k(n_d + r(x)) + 1\} \;=\; O\left(\frac{k\sigma\log(\Lambda\sigma)}{\log\lambda}\right),$$
$$\phi_d\lambda^{r(x)} \;\le\; \lambda^{1/2}(\Lambda\sigma)^{G/2} \quad\text{and}\quad \phi_d^2\lambda^{r(x)+r'(x)} \;\le\; (\Lambda\sigma)^G.$$

Hence it follows from Theorem 5.7 and (4.10) that

$$\begin{aligned}
&|\mathbf{P}[D_d > 2n_d + x] - \mathbf{P}[D_d^* > 2n_d + x]| \\
&\le \{\eta_1'(r(x), r'(x)) + \eta_2'(r(x), r'(x))\}(\Lambda\sigma)^{-1/2} \\
&\qquad + \eta_3'(r(x), r'(x))(\Lambda\sigma)^{-1} + \eta_4'(r(x), r'(x))(\Lambda\sigma)^{-\gamma/2} \\
&= O\Bigg(\lambda^{1/2}\left(\frac{k\sigma\log(\Lambda\sigma)}{\log\lambda}\right)^2 (\Lambda\sigma)^{-(1-3G)/2}\log(\Lambda\sigma) \\
&\qquad + \lambda^{1/2}\left(\frac{k\sigma\log(\Lambda\sigma)}{\log\lambda}\right)(\Lambda\sigma)^{-(1-G)/2}\log(\Lambda\sigma) \\
&\qquad + \lambda^2\sigma(\Lambda\sigma)^{G-1} + \lambda^2(\Lambda\sigma)^{-1} \\
&\qquad + \lambda^{(5+\gamma)/2}\left(\frac{\log(\Lambda\sigma)}{\log\lambda}\right)^{1/2}(\Lambda\sigma)^{-(\gamma - G(2-\gamma))/2}\Bigg).
\end{aligned}$$

Noting that $\lambda = O(1 + \sigma + \sqrt{k\sigma})$ from (4.9), that $k\sigma$ is bounded by some fixed ρ_0, and remembering that $\gamma \le 1/2$, the final term is of largest asymptotic order. Also, from Lemma 5.8, recalling from (4.17) that $\mathbf{P}[D_d^* > 2n_d + x] = \mathbf{E}\exp\left\{-\theta W_{k,\sigma} W_{k,\sigma}'\right\}$ with $\theta = \frac{\lambda^2}{(\lambda-\lambda_2)}(\lambda - \sigma)\phi_d^2\lambda^x$, and taking $x = \left\lceil \frac{G\log(\Lambda\sigma) - 2\log\phi_d}{\log\lambda} \right\rceil$, we have the upper tail estimate

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

Comparing the exponents of $\Lambda\sigma$, the best choice of G is $G = \gamma/(4-\gamma)$, which makes $G = (\gamma - G(2-\gamma))/2$, proving the theorem. □

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 5.9 matches Theorem 3.10

closely for $\rho \leq 1$, 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 (6.4),

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

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 3.11 could for instance also be proved.

5.3 The distance to the continuous circle – continuous time model

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, given in (5.23), satisfy the Poincaré equation (5.24). The argument runs parallel to that in Section 3.4, though it is rather more complicated. Define the operator Ξ, analogous to the operator Ψ in Section 3.4, by

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

Let

$$\mathcal{G} \ := \ \left\{ \gamma = (\gamma_1, \gamma_2) : [0,\infty)^2 \to [0,1] : \|\gamma\|_{\mathcal{G}} := \sup_{\theta>0} \max \left\{ \frac{|\gamma_1(\theta)|, |\gamma_2(\theta)|}{\theta^2} \right\} < \infty \right\},$$

and

$$\mathcal{H} := \left\{ \psi = (\psi_1, \psi_2) : [0,\infty)^2 \to [0,1] : (\psi_1(\theta) - \{1-\theta\}, \psi_2(\theta) - \{1 - \theta(\lambda - \sigma)\}) \in \mathcal{G} \right\}.$$

Then $\mathcal{H}$ contains $\varphi_{k,\sigma} = (\varphi_1, \varphi_2)$ as defined in (5.23), since

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

and taking limits in (5.8) shows that $\operatorname{Var} W_{k,\sigma}$ exists. Furthermore, from (5.24), Ξ has $\varphi_{k,\sigma}$ as a fixed point. We next show that Ξ is a contraction on $\mathcal{H}$.

Lemma 5.10 *The operator Ξ is a contraction on $\mathcal{H}$, and, for all $\psi, \chi \in \mathcal{H}$,*

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

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

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

so that

$$\begin{aligned} \frac{|(\Xi\psi)_1(\theta) - (\Xi\chi)_1(\theta)|}{\theta^2} \ &\le \ \frac{1}{\lambda^2} \frac{\left|\psi_2\left(\frac{\theta}{\lambda}\right) - \chi_2\left(\frac{\theta}{\lambda}\right)\right|}{\left(\frac{\theta}{\lambda}\right)^2} \\ &\quad + \frac{\sigma}{\lambda^2} \frac{\left|\psi_1\left(\frac{\theta}{\lambda}\right) - \chi_1\left(\frac{\theta}{\lambda}\right)\right|}{\left(\frac{\theta}{\lambda}\right)^2} \\ &\le \ \frac{\sigma + 1}{\lambda^2} \|\psi - \chi\|_{\mathcal{G}}. \end{aligned} \tag{5.25}$$

Similarly,

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

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 $\varphi_{k,\sigma} = (\varphi^{(1)}, \varphi^{(2)})$ given in (5.23), we have

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

so that

$$\|\varphi_{k,\sigma} - \psi\|_{\mathcal{G}} \leq \frac{\lambda^2}{\lambda^2 - (2k\sigma+1)} \|\Xi\psi - \psi\|_{\mathcal{G}}. \tag{5.26}$$

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

$$\begin{aligned} \psi_1(\theta) &= \frac{1}{1+\theta}, \\ \psi_2(\theta) &= \frac{1}{1+\theta(\lambda-\sigma)}; \end{aligned} \tag{5.27}$$

Lemma 5.8 shows that this pair dominates $\phi_{k,\sigma}$.

Lemma 5.11 *For ψ given in (5.27), 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

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

where

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

Moreover,

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

From Taylor's expansion, it follows that

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

Hence

$$
\frac{|(\Xi\psi)_1(\theta)-\psi_1(\theta)|}{\theta^2} \le \frac{\sigma(2(\lambda-1)+\sigma)}{2\lambda^2}. \tag{5.28}
$$

Since

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

this is enough for the first component.

Similarly,

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

where

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

Using (4.4), we obtain

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

From Taylor's expansion, it now follows that

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

Hence

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

completing the proof. $\square$

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

Theorem 5.12 *As $k\sigma \to 0$,* $\quad \mathcal{L}(W_{k,\sigma}) \;\to\; \mathrm{NE}(1)$.

Proof: Let $\varphi_{k,\sigma}$ be as in (5.23), and ψ as in (5.27). Then $(\varphi_{k,\sigma})_1$ is the Laplace transform of

$$\mathcal{L}(\sqrt{\lambda-\sigma}\,W_{k,\sigma}) \;:=\; \mathcal{L}(\sqrt{\lambda-\sigma}\,W_{k,\sigma}\,|\,\hat{M}(0)=\mathbf{e}^{(1)}),$$

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

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

However, using Lemma 5.11 and (5.26), we obtain

$$\begin{aligned}
&\|\varphi_{k,\sigma}-\psi\|_{\mathcal{G}} \\
&\quad\le\; \left(\frac{\lambda^2}{\lambda^2-1-2k\sigma}\right)\|\Xi\psi-\psi\|_{\mathcal{G}} \\
&\quad\le\; 2k\sigma\,\frac{\lambda^2-\lambda\sigma-1+k\sigma}{\lambda^2-1-2k\sigma}.
\end{aligned}$$

Now, from (4.4), we have

$$\begin{aligned}
\frac{\lambda^2-\lambda\sigma-1+k\sigma}{\lambda^2-1-2k\sigma} &= \frac{\lambda-1+(3k-1)\sigma}{(\lambda-1)(1+\sigma)} \\
&= \frac{1}{1+\sigma}\left\{1+\frac{(3k-1)\sigma}{2k\sigma}\,(\lambda-\sigma)\right\} \\
&\le 1+\tfrac{3}{2}(1+2k\sigma),
\end{aligned}$$

this last from (4.8). Hence

$$\|\varphi_{k,\sigma}-\psi\|_{\mathcal{G}} \;\le\; k\sigma(5+6k\sigma) \;\to\; 0, \tag{5.29}$$

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

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.

Lemma 5.13 *Let W, W' be independent NE(1) variables. Then, for all $\theta > 0$,*

$$\left|\mathbf{E}\exp\left\{-\theta(\lambda-\sigma)W_{k,\sigma}W'_{k,\sigma}\right\} - \mathbf{E}e^{-\theta WW'}\right| \;\le\; \theta^2 k\sigma\left\{25+24k\sigma\right\}.$$

Proof: As in the proof of Lemma 3.15, with $\varphi_{k,\sigma}$ as in (5.23) and ψ as in (5.27), we have

$$\begin{aligned}
&\mathbf{E}\exp\left\{-\theta(\lambda-\sigma)W_{k,\sigma}W'_{k,\sigma}\right\} - \mathbf{E}e^{-\theta WW'} \\
&\quad= \mathbf{E}\left\{\varphi^{(1)}_{k,\sigma}(\theta\sqrt{\lambda-\sigma}\,W'_{k,\sigma})\right\} - \mathbf{E}\psi_1(\theta W) \\
&\quad= \mathbf{E}\left\{(\Xi\varphi_{k,\sigma})_1(\theta\sqrt{\lambda-\sigma}\,W'_{k,\sigma})\right\} - \mathbf{E}\left\{(\Xi\psi)_1(\theta\sqrt{\lambda-\sigma}\,W'_{k,\sigma})\right\} \\
&\quad\quad+\mathbf{E}\left\{(\Xi\psi)_1(\theta\sqrt{\lambda-\sigma}\,W'_{k,\sigma})\right\} - \mathbf{E}\left\{\psi_1(\theta\sqrt{\lambda-\sigma}\,W'_{k,\sigma})\right\} \\
&\quad\quad+\mathbf{E}\left\{\psi_1(\theta\sqrt{\lambda-\sigma}\,W'_{k,\sigma})\right\} - \mathbf{E}\psi_1(\theta W).
\end{aligned} \tag{5.30}$$

Now (5.25) in the proof of Lemma 5.10 gives

$$\begin{aligned}
&\left|\mathbf{E}\left\{(\Xi\varphi_{k,\sigma})_1(\theta\sqrt{\lambda-\sigma}\,W'_{k,\sigma})\right\} - \mathbf{E}\left\{(\Xi\psi)_1(\theta\sqrt{\lambda-\sigma}\,W'_{k,\sigma})\right\}\right| \\
&\quad \le\; \theta^2(\lambda-\sigma)\,\frac{\sigma+1}{\lambda^2}\,\|\varphi_{k,\sigma}-\psi\|_{\mathcal{G}}\mathbf{E}\{(W'_{k,\sigma})^2\}.
\end{aligned} \tag{5.31}$$

Since, from (4.13), (4.16), (5.8) and (4.10),

$$\mathbf{E}\{(W'_{k,\sigma})^2\} \;\le\; (1+\omega^2)/(\lambda-\sigma) \;\le\; 2/(\lambda-\sigma),$$

and since $\lambda^2 > \lambda > \sigma+1$ from (4.5), it follows from (5.29) that the expression (5.31) can be bounded by $2\theta^2 k\sigma(5+6k\sigma)$. Similarly, from (5.28) in the proof of Lemma 5.11,

$$\begin{aligned}
&\left|\mathbf{E}\left\{(\Xi\psi)_1(\theta\sqrt{\lambda-\sigma}\,W'_{k,\sigma})\right\} - \mathbf{E}\left\{\psi_1(\theta\sqrt{\lambda-\sigma}\,W'_{k,\sigma})\right\}\right| \\
&\quad \le\; 2\theta^2\,\frac{\sigma\{2(\lambda-1)+\sigma\}}{2\lambda^2} \;\le\; \theta^2\,\frac{\sigma^2(4k+1)}{1+\sigma} \;\le\; 5\theta^2 k\sigma,
\end{aligned}$$

from (4.8) and because $\lambda^2 > 1+\sigma$. Then, with $W \sim \mathrm{NE}\,(1)$ independent of $W'_{k,\sigma}$, we have

$$\mathbf{E}\left\{\psi_1(\theta\sqrt{\lambda-\sigma}\,W'_{k,\sigma})\right\} \;=\; \mathbf{E}\left\{e^{-\theta\sqrt{\lambda-\sigma}\,WW'_{k,\sigma}}\right\} \;=\; \mathbf{E}\left\{\varphi^{(1)}_{k,\sigma}(\theta W)\right\},$$

and hence, from (5.29) in the proof of Theorem 5.12, it follows that

$$\begin{aligned}
\left|\mathbf{E}\left\{\psi_1(\theta\sqrt{\lambda-\sigma}\,W'_{k,\sigma})\right\} - \mathbf{E}\psi_1(\theta W)\right| \;&=\; \left|\mathbf{E}\left\{\varphi^{(1)}_{k,\sigma}(\theta W)\right\} - \mathbf{E}\psi_1(\theta W)\right| \\
&\le\; 2\theta^2 k\sigma(5+6k\sigma).
\end{aligned}$$

The assertion now follows from (5.30). □

Recalling from (3.24) that

$$\mathbf{E}e^{-\theta WW'} \;=\; \int_0^\infty \frac{e^{-y}}{1+\theta y}\,dy, \tag{5.32}$$

we can now derive the analogue of Theorem 3.16.

Theorem 5.14 *Let T denote a random variable on $\mathbf{R}$ with distribution given by*

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

Then

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

uniformly in $k\sigma \le 1/2$, where $\mathcal{L}(D^*_d)$ is as in (4.17).

Remark. For comparison with Theorem 3.16, note that $\frac{\lambda-1}{2} = \rho(1+O(\rho))$ as $\rho = k\sigma \to 0$, so that $\rho(D_d^* - 2n_d)$ can be approximated by T to the same order of accuracy.

Proof: We use an argument similar to the proof of Theorem 3.16. Putting

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

we have, as before,

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

we also write

$$\beta(\lambda) \;:=\; \frac{\lambda^2\phi_d^2}{\lambda-\lambda_2}.$$

Then, because $\lambda^{-1} \le \phi_d \le 1$ and $\lambda_2 < 0$, and from (4.9), we have

$$\beta(\lambda) \;\le\; \lambda \;\le\; 1+\sigma+\sqrt{\sigma(2k-1)}$$

and

$$\begin{aligned}\beta(\lambda) \;&\ge\; (\lambda-\lambda_2)^{-1} \;=\; \frac{1}{2\lambda-1-\sigma} \\ &\ge\; 1 - \frac{\sigma+2\sqrt{\sigma(2k-1)}}{1+\sigma+2\sqrt{\sigma(2k-1)}} \;\ge\; 1-\{\sigma+2\sqrt{\sigma(2k-1)}\}.\end{aligned}$$

This in turn gives

$$\begin{aligned}|\beta(\lambda)-1| \;&\le\; \sigma+2\sqrt{\sigma(2k-1)} \;=:\; \Gamma(\sigma,k) \\ &=\; O\left(\sigma+\sqrt{k\sigma}\right).\end{aligned}$$

For the main part of the distribution, writing $\Delta_d = D_d^* - 2n_d$, we have

$$\begin{aligned}&\mathbf{P}\left[\tfrac{\lambda-1}{2}\Delta_d > z\right] - \mathbf{P}[T>z] \\ &\quad= \mathbf{P}\left[\tfrac{\lambda-1}{2}\Delta_d > z\right] - \int_0^\infty e^{-y}\left(1+y\beta(\lambda)e^{2zc(\lambda)}\right)^{-1}dy \qquad (5.33)\\ &\qquad + \int_0^\infty e^{-y}\left(1+y\beta(\lambda)e^{2zc(\lambda)}\right)^{-1}dy - \mathbf{P}[T>zc(\lambda)] \qquad (5.34)\\ &\qquad + \mathbf{P}[T>zc(\lambda)] - \mathbf{P}[T>z]. \qquad (5.35)\end{aligned}$$

Now, considering first (5.33), we combine (5.32), Lemma 5.13 and (4.17) to yield

$$\begin{aligned}&\left|\mathbf{P}\left[\tfrac{\lambda-1}{2}\Delta_d > z\right] - \int_0^\infty e^{-y}\left(1+y\beta(\lambda)e^{2zc(\lambda)}\right)^{-1}\,dy\right| \\ &= \left|\mathbf{E}\exp\left\{-\beta(\lambda)(\lambda-\sigma)e^{2zc(\lambda)}W_{k,\sigma}W'_{k,\sigma}\right\} - \mathbf{E}\exp\left\{-\beta(\lambda)e^{2zc(\lambda)}WW'\right\}\right| \\ &\le \beta(\lambda)^2e^{4zc(\lambda)}k\sigma\left\{25+24k\sigma\right\} \;\le\; \lambda^2e^{4z}k\sigma\left\{25+24k\sigma\right\}.\end{aligned}$$

With (3.26), we have, for (5.34), that

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

Similarly, for (5.35), it follows that

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

Now, for $z > 0$, because $0 \le 1 - c(\lambda) \le \frac{\lambda-1}{2} \le k\sigma$ and from Taylor's expansion, this gives

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

if $z \le 0$, we have

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

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

$$\begin{aligned}
&\mathbf{P}\left[\tfrac{\lambda-1}{2}\Delta_d > z\right] - \mathbf{P}[T > z] \\
&\quad \le\ k\sigma e^{4z}\lambda^2 \{25 + 24k\sigma\} + \Gamma(\sigma, k) + 2k\sigma|z| \min\{1, e^{2z(1-k\sigma)}\} \\
&\quad \le\ C_1 \left\{ k\sigma(e^{4z} + 1) + \sqrt{k\sigma} \right\}, \qquad (5.36)
\end{aligned}$$

for some constant C_1.

For the large values of z, where the bound given in (5.36) becomes useless, we can estimate the upper tails of the random variables separately. First, for $x \in \mathbf{Z}$, we have

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

so that, by Lemma 5.8, it follows that

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

For the upper tail of T, from (3.29), we have

$$\mathbf{P}[T > z] \;\le\; e^{-2z}\log(1+e^{2z}) \;\le\; (1+2z)e^{-2z},$$

so that, combining these two tail estimates,

$$\left|\mathbf{P}\left[\tfrac{\lambda-1}{2}\Delta_d > z\right] - \mathbf{P}[T > z]\right| \;\le\; C_2(1+z)e^{-2z(1-\sigma k)}, \tag{5.37}$$

in $z > 0$, uniformly in $k\sigma \le 1/2$, for some constant $C_2 > 0$. Applying the bound (5.36) when $z \le (6-2k\sigma)^{-1}\log(1/k\sigma)$ and (5.37) for all larger z, and remembering that T has bounded density, so that the discrete nature of D_d^* requires only a small enough correction, a bound of the required order follows. □

6 Appendix: proof of Lemma 5.3

In this section, we make frequent use of the inequality (4.10). From (4.15) and (4.14) 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 (5.6); for (5.10), use $\sigma+1-\lambda=\lambda_2$ and $\sigma+1-\lambda_2=\lambda$, from (4.7). Then, using the same equations, we have

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

and (5.7) follows because $a^{n+1}+b^{n+1} \le a^n(a+b)$ whenever $0 \le b \le a$, and because $\lambda \ge 1$.

Now define

$$X(n) \;:=\; \hat{M}^{(1)}(n) - \sigma\{\hat{M}^{(1)}(n-1) + 2k\hat{M}^{(2)}(n-1)\}, \quad n \ge 1, \tag{6.1}$$

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,

$$\begin{aligned}
&W^{(i)}(n+1) - W^{(i)}(n)\\
&\quad= \lambda_i^{-n-1}f_1^{(i)}\{\hat{M}^{(1)}(n+1) + (\lambda_i-\sigma)\hat{M}^{(2)}(n+1)\\
&\qquad\qquad - \lambda_i\hat{M}^{(1)}(n) - \lambda_i(\lambda_i-\sigma)\hat{M}^{(2)}(n)\}\\
&\quad= \lambda_i^{-n-1}f_1^{(i)}\{\hat{M}^{(1)}(n+1) - \sigma\hat{M}^{(1)}(n) - 2k\sigma\hat{M}^{(2)}(n)\}\\
&\quad= \lambda_i^{-n-1}f_1^{(i)}X(n+1),
\end{aligned} \tag{6.2}$$

where we have used $(\lambda_i - 1)(\lambda_i - \sigma) = 2k\sigma$, from (4.4), and the branching recursion.

Since

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

and because, if $X \sim \mathrm{Bi}\,(n,p)$, then $\mathbf{E}\{(X-np)^3\} = np(1-p)((1-p)^2 - p^2)$, it follows that

$$
\begin{aligned}
\mathbf{E}\{X^3(n+1) \,|\, \mathcal{F}(n)\} &\le \frac{\sigma}{\Lambda}\left(1 - \frac{\sigma}{\Lambda}\right)^3\{\hat{M}^{(1)}(n) + 2k\hat{M}^{(2)}(n)\}\Lambda \\
&\le \sigma\{\hat{M}^{(1)}(n) + 2k\hat{M}^{(2)}(n)\}
\end{aligned}
$$

also, we have

$$
\mathbf{E}X^2(n+1) \;\le\; 2k\sigma\lambda^n \quad \text{and} \quad \mathbf{E}X^3(n+1) \;\le\; 2k\sigma\lambda^n,
$$

from (5.7). Thus, immediately,

$$
\mathbf{E}\{(W^{(i)}(n+1) - W^{(i)}(n))^2\} \;\le\; 2k\sigma(f_1^{(i)})^2\lambda_i^{-2n-2}\lambda^n. \tag{6.3}
$$

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

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

and the formulae (5.8) and (5.9) follow.

Moreover, from (4.7) and (4.15), and then from (6.2) and (4.13), we have

$$
\begin{aligned}
(\lambda - \lambda_2)\hat{M}^+(n) &= \left(\frac{\lambda^{n+1}}{f_1^{(1)}}W^{(1)}(n) - \frac{\lambda_2^{n+1}}{f_1^{(2)}}W^{(2)}(n)\right) \\
&= (\lambda^{n+1} - \lambda_2^{n+1}) + \sum_{j=1}^{n} c_{jn}X(j),
\end{aligned} \tag{6.4}
$$

where

$$
0 \;\le\; c_{jn} \;:=\; \lambda^{n+1-j} - \lambda_2^{n+1-j} \;\le\; 2\lambda^{n+1-j}. \tag{6.5}
$$

Hence

$$
\begin{aligned}
(\lambda - \lambda_2)^2\mathrm{Var}\,\hat{M}^+(n) &= \sum_{j=1}^{n} c_{jn}^2\,\mathrm{Var}\,X(j) \;\le\; 8k\sigma\sum_{j=1}^{n}\lambda^{2n-(j-1)} \\
&\le 8k\sigma\lambda^{2n+1}/(\lambda - 1) \;=\; 4\omega^2\lambda^{2n+2},
\end{aligned}
$$

and

$$
\begin{aligned}
(\lambda-\lambda_2)^3\mathbf{E}\{(\hat{M}^+(n)-\mathbf{E}\hat{M}^+(n))^3\} \;&=\; \sum_{j=1}^{n} c_{jn}^3\mathbf{E}X^3(j) \\
\le\; 16k\sigma\sum_{j=1}^{n}\lambda^{3n-2(j-1)} \;&\le\; 16k\sigma\lambda^{3n+2}/(\lambda^2-1) \;=\; 8\omega^2\lambda^{3n+2}.
\end{aligned}
$$

(5.11) is now immediate, using (5.10); furthermore,

$$
\begin{aligned}
&\mathbf{E}\{(\hat{M}^+(n))^3\} \\
&= \mathbf{E}\{(\hat{M}^+(n)-\mathbf{E}\hat{M}^+(n))^3\} + 3\mathbf{E}\hat{M}^+(n)\,\mathrm{Var}\,\hat{M}^+(n) + \{\mathbf{E}\hat{M}^+(n)\}^3 \\
&\le 8\omega^2\lambda^{3n-1} + 24\omega^2\lambda^{3n} + 8\lambda^{3n} \;\le\; 40\lambda^{3n},
\end{aligned}
$$

and the bounds in (5.12) and (5.13) follow because $\log x \le n\log\lambda + \lambda^{-n}x$ in $x \ge 0$. □

Acknowledgement. We would like to thank Svante Janson for very helpful discussions concerning the comparison with [13], and the referee for contributing a large number of extremely useful suggestions. We also thank the Institute for Mathematical Sciences at the National University of Singapore for their hospitality while part of the work was accomplished.

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, Basel.

[3] Athreya, K.B. and Ney, P.E. (1972). *Branching Processes.* Springer, 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, Mass.

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

[7] Barbour, A.D. and Reinert, G. (2001). Small Worlds. *Random Structures and Algorithms* **19**, 54–74. Correction: *ibid* **25**, 115 (2004).

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

[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] Janson, S. (1999) One, two and three times $\log n/n$ for paths in a complete graph with random weights. *Combin. Probab. Comput.* **8**, 347–361.

[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.