CrossMark

Escape regimes of biased random walks on Galton–Watson trees

Adam Bowditch[1] (ID)

Abstract We study biased random walk on subcritical and supercritical Galton–Watson trees conditioned to survive in the transient, sub-ballistic regime. By considering offspring laws with infinite variance, we extend previously known results for the walk on the supercritical tree and observe new trapping phenomena for the walk on the subcritical tree which, in this case, always yield sub-ballisticity. This is contrary to the walk on the supercritical tree which always has some ballistic phase.

Keywords Random walk in random environment · Galton–Watson tree · Infinite variance · Infinitely divisible distributions · Sub-ballistic

Mathematics Subject Classification Primary 60K37 · 60F05; Secondary 60E07 · 60J80

1 Introduction

In this paper, we investigate biased random walks on subcritical and supercritical Galton–Watson trees. These are a natural setting for studying trapping phenomena as dead-ends, caused by leaves in the trees, slow the walk. These models can be used to approach more difficult problems concerning biased random walks on percolation clusters (as studied in [11,13,24]) and random walk in random environment (see for example [18,25]). For a recent review of trapping phenomena and random walk in random environment we direct the reader to [3] which covers recent developments in

✉ Adam Bowditch
a.bowditch@warwick.ac.uk

[1] University of Warwick, Coventry, UK

a range of models of directionally transient and reversible random walks on underlying graphs such as supercritical GW-trees and supercritical percolation clusters.

For supercritical GW-trees with leaves, it has been shown in [21] that, for a suitably large bias away from the root, the dead-ends in the environment create a sub-ballistic regime. In this case, it has further been observed in [4], that if the offspring distribution has finite variance then the walker follows a polynomial escape regime but cannot be rescaled properly due to a certain lattice effect. (In [5,15] it is shown that, in a related model where the conductance along each is chosen randomly according to a distribution satisfying a certain non-lattice assumption, the tail of the trapping time obeys a pure power law and the rescaled walk converges in distribution.) Here we show that, when the offspring law has finite variance, the walk on the subcritical GW-tree conditioned to survive experiences similar trapping behaviour to the walk on the supercritical GW-tree shown in [4]. However, the main focus of the article concerns offspring laws belonging to the domain of attraction of some stable law with index $\alpha \in (1, 2)$. In this setting, although the distribution of time spent in individual traps has polynomial tail decay in both cases, the exponent varies with α in the subcritical case and not in the supercritical case. This results in a polynomial escape of the walk which is always sub-ballistic in the subcritical case unlike the supercritical case which always has some ballistic phase.

We now describe the model of a biased random walk on a subcritical GW-tree conditioned to survive which will be the main focus of the article. Let $f(s) := \sum_{k=0}^{\infty} p_k s^k$ denote the probability generating function of the offspring law of a GW-process with mean $\mu > 0$ and variance $\sigma^2 > 0$ (possibly infinite) and let Z_n denote the nth generation size of a process with this law started from a single individual, i.e. $Z_0 = 1$. Such a process gives rise to a random tree where individuals in the process are represented by vertices and undirected edges connect individuals with their offspring.

For a fixed tree T let ρ denote its root, Z_n^T the size of the nth generation, $\overleftarrow{x}$ the parent of $x \in T$, $c(x)$ the set of children of x, $d_x := |c(x)|$ the out-degree of x, $d(x, y)$ the graph distance between vertices x, y, $|x| := d(\rho, x)$ the graph distance between x and the root and T_x to be the descendent tree of x. A β-biased random walk on a fixed, rooted tree T is a random walk $(X_n)_{n \geq 0}$ on T which is β-times more likely to make a transition to a given child of the current vertex than the parent (which are the only options). That is, the random walk is the Markov chain started from $X_0 = z$ defined by the transition probabilities

$$P_z^T(X_{n+1} = y | X_n = x) := \begin{cases} \frac{1}{1+\beta d_x} & \text{if } y = \overleftarrow{x}, \\ \frac{\beta}{1+\beta d_x}, & \text{if } y \in c(x), \ x \neq \rho, \\ \frac{1}{d_\rho}, & \text{if } y \in c(x), \ x = \rho, \\ 0, & \text{otherwise.} \end{cases}$$

We use $\mathbb{P}_\rho(\cdot) := \int P_\rho^T(\cdot) \mathbf{P}(dT)$ for the annealed law obtained by averaging the quenched law P_ρ^T over a law $\mathbf{P}$ on random trees with a fixed root ρ. In general we will drop the superscript T and subscript ρ when it is clear to which tree we are referring and we start the walk at the root.

We will mainly be interested in GW-trees $\mathcal{T}$ which survive, that is $\mathcal{H}(\mathcal{T}) := \sup\{n \geq 0 : Z_n > 0\} = \infty$. It is classical (e.g. [2]) that when $\mu > 1$ there is some strictly positive probability $1 - q$ that $\mathcal{H}(\mathcal{T}) = \infty$ whereas when $\mu \leq 1$ we have that $\mathcal{H}(\mathcal{T})$ is almost surely finite. However, it has been shown in [17] that there is some well defined probability measure $\mathbf{P}$ over f-GW trees conditioned to survive for infinitely many generations which arises as a limit of probability measures over f-GW trees conditioned to survive at least n generations. Henceforth, we assume $\mathbf{P}$ is this law and X_n is a random walk on an f-GW-tree conditioned to survive.

The main object of interest is $|X_n|$, that is, how the distance from the root changes over time. Due to the typical size of finite branches in the tree being small and the walk not backtracking too far we shall see that $|X_n|$ has a strong inverse relationship with the first hitting times $\Delta_n := \inf\{m \geq 0 : X_m \in \mathcal{Y}, |X_m| = n\}$ of levels along the backbone $\mathcal{Y} := \{x \in \mathcal{T} : \mathcal{H}(\mathcal{T}_x) = \infty\}$ so for much of the paper we will consider this instead. It will be convenient to consider the walk as a trapping model. To this end we define the underlying walk $(Y_k)_{k \geq 0}$ defined by $Y_k := X_{\eta_k}$ where $\eta_0 := 0$ and $\eta_k := \inf\{m > \eta_{k-1} : X_m, X_{m-1} \in \mathcal{Y}\}$ for $k \geq 1$.

When X_n is a walk on an f-GW tree conditioned to survive for f supercritical ($\mu > 1$), it has been shown in [21] that $\nu(\beta) := \lim_{n \to \infty} |X_n|/n$ exists $\mathbb{P}$-a.s. and is positive if and only if $\mu^{-1} < \beta < f'(q)^{-1}$ in which case we call the walk ballistic. Furthermore, although no explicit expression for the speed ν is known, a description of the invariant distribution of the environment seen from the particle is used in [1] to give an expression of the speed in terms of the annealed expectation. This expression coincides with the speed of the walk on a certain regular tree where each vertex has some number of children m_β; in particular, it can be seen that $m_\beta \leq \mu$ therefore the randomness of the tree slows the walk. If $\beta \leq \mu^{-1}$ then the walk is recurrent because the average drift of Y acts towards the root. When $\beta \geq f'(q)^{-1}$ the walker expects to spend an infinite amount of time in the finite trees which hang off $\mathcal{Y}$ (see Fig. 5 in Sect. 10) thus causing a slowing effect which results in the walk being sub-ballistic. In this case, the correct scaling for some non-trivial limit is n^γ where γ will be defined later in (1.1). In particular, it has been shown in [4] that, when $\sigma^2 < \infty$, the laws of $|X_n|n^{-\gamma}$ are tight and, although $|X_n|n^{-\gamma}$ doesn't converge in distribution, we have that $\Delta_n n^{-1/\gamma}$ converges in distribution under $\mathbb{P}$ along certain subsequences to some infinitely divisible law. In Sect. 10 we extend this result by relaxing the condition that the offspring law has finite variance and instead requiring only that it belongs to the domain of attraction of some stable law of index $\alpha > 1$.

Recall that the offspring law of the process is given by $\mathbf{P}(\xi = k) := p_k$, then we define the size-biased distribution by the probabilities $\mathbf{P}(\xi^* = k) := kp_k\mu^{-1}$. It can be seen (e.g. [16]) that the subcritical ($\mu < 1$) GW-tree conditioned to survive coincides with the following construction: Starting with a single special vertex, at each generation let every normal vertex give birth onto normal vertices according to independent copies of the original offspring distribution and every special vertex give birth onto vertices according to independent copies of the size-biased distribution, one of which is chosen uniformly at random to be special. Unlike the supercritical tree which has infinitely many infinite paths, the backbone of the subcritical tree conditioned to survive consists of a unique semi-infinite path from the initial vertex ρ. We call the vertices not on $\mathcal{Y}$ which are children of vertices on $\mathcal{Y}$ buds and the finite

trees rooted at the buds traps (see Fig. 2 in Sect. 3). In this paper we consider walks with positive bias therefore the walk is transient and only returns to the starting vertex ρ finitely often. Moreover, we are interested in the case where the trapping times are heavy tailed and therefore since the traps are i.i.d. the walk closely resembles a one dimensional directed trap model as studied in [26].

Briefly, the phenomena that can occur in the subcritical case are as follows. When $\mathbf{E}[\xi \log^+(\xi)] < \infty$ and $\mu < 1$ there exists a limiting speed $\nu(\beta)$ such that $|X_n|/n$ converges almost surely to $\nu(\beta)$ under $\mathbb{P}$; moreover, the walk is ballistic ($\nu(\beta) > 0$) if and only if $1 < \beta < \mu^{-1}$ and $\sigma^2 < \infty$. This essentially follows from the argument used in [21] (to show the corresponding result on the supercritical tree) with the fact that, by (2.1) and (5.2), the conditions given are precisely the assumptions needed so that the expected time spent in a branch is finite (see [8] or [9] for further detail). The sub-ballistic regime has four distinct phases. When $\beta \leq 1$ the walk is recurrent and we are not concerned with this case here. When $1 < \beta < \mu^{-1}$ and $\sigma^2 = \infty$ the expected time spent in a trap is finite and the slowing of the walk is due to the large number of buds. When $\beta\mu > 1$ and $\sigma^2 < \infty$, the expected time spent in a subcritical GW-tree forming a trap is infinite because the strong bias forces the walk deep into traps and long sequences of movements against the bias are required to escape. In the final case for the subcritical tree ($\beta\mu > 1$, $\sigma^2 = \infty$) slowing effects are caused by both strong bias and the large number of buds.

Figure 1 is the phase diagram for the almost sure limit of $\log(|X_n|)/\log(n)$ (which is the leading order polynomial exponent in the scaling of $|X_n|$ relative to β and μ) where the offspring law has stability index α (which is 2 when $\sigma^2 < \infty$) and we define
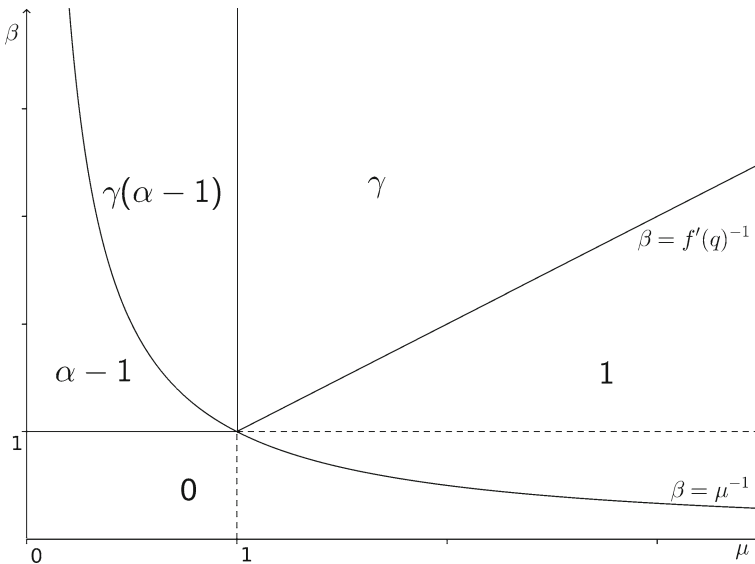

Fig. 1 Phase diagram for the leading order polynomial exponent in the scaling of the walk relative to the mean of the offspring law and bias of the walk

$$\gamma := \begin{cases} \frac{\log(f'(q)^{-1})}{\log(\beta)}, & \mu > 1, \\ \frac{\log(\mu^{-1})}{\log(\beta)}, & \mu < 1, \end{cases} \tag{1.1}$$

where we note that $f'(q)$ and μ are the mean number of offspring from vertices in traps of the supercritical and subcritical trees respectively. Strictly, $f'(q)$ isn't a function of μ therefore the line $\beta = f'(q)^{-1}$ is not well defined; Fig. 1 shows the particular case when the offspring distribution belongs to the geometric family. It is always the case that $f'(q) < 1$ therefore some such region always exists however the parametrisation depends on the family of distributions.

When the offspring law has finite variance, the limiting behaviour of $|X_n|$ on the supercritical and subcritical trees is very similar. Both have a regime with linear scaling (which is, in fact, almost sure convergence of $|X_n|/n$) and a regime with polynomial scaling caused by the same phenomenon of deep traps (which results in $|X_n|n^{-\gamma}$ not converging). When the offspring law has infinite variance, the bud distribution of the subcritical tree has infinite mean which causes an extra slowing effect which isn't seen with the supercritical tree. This equates for the different exponents observed in the two models as shown in Fig. 1. The walk on the critical ($\mu = 1$) tree experiences a similar trapping mechanism to the subcritical tree; however, the slowing is more extreme and belongs to a different universality class which had been shown in [10] to yield a logarithmic escape rate.

2 Statement of main theorems and proof outline

In this section we introduce the three sub-ballistic regimes in the subcritical case and the one further regime for the infinite variance supercritical case that we consider here. We then state the main theorems of the paper.

The subcritical tree has bud distribution $\xi^* - 1$ where $\mathbf{P}(\xi^* = k) = kp_k\mu^{-1}$ which yields the following important property relating the size biased and offspring distributions

$$\mathbf{E}\left[\varphi(\xi^*)\right] = \sum_{k=1}^{\infty} \varphi(k)\frac{kp_k}{\mu} = \mathbf{E}\left[\varphi(\xi)\xi\right]\mu^{-1}. \tag{2.1}$$

Choosing φ to be the identity we have finite mean of the size-biased distribution if and only if the variance of the offspring distribution is finite. This causes a phase transition for the walk that isn't seen in the supercritical tree. The reason for this is that in the corresponding decomposition for the supercritical tree we have subcritical GW-trees as leaves but the number of buds is exponentially tilted and therefore maintains moment properties.

If the offspring law belongs to the domain of attraction of some stable law of index $\alpha \in (1, 2)$ then taking $\varphi(x) = x\mathbf{1}_{\{x \le t\}}$ shows that the size biased distribution belongs to the domain of attraction of some stable law with index $\alpha - 1$ and allows us to attain properties of the scaling sequences (see for example [12, IX.8]).

The first case we consider is when $\beta\mu < 1$ but $\sigma^2 = \infty$; we refer to this as the infinite variance, finite excursion case:

Definition 1 (*IVFE*) The offspring distribution has mean μ satisfying $1 < \beta < \mu^{-1}$ and belongs to the domain of attraction of a stable law of index $\alpha \in (1, 2)$.

Under this assumption we let L vary slowly at ∞ such that as $x \to \infty$

$$\mathbb{E}\left[\xi^2 \mathbf{1}_{\{\xi \leq x\}}\right] \sim x^{2-\alpha} L(x) \tag{2.2}$$

and choose $(a_n)_{n\geq 1}$ to be some scaling sequence for the size-biased law such that for any $x > 0$, as $n \to \infty$ we have $\mathbf{P}(\xi^* \geq x a_n) \sim x^{-(\alpha-1)} n^{-1}$. Moreover for some slowly varying function $\tilde{L}$ we have that $a_n = n^{\frac{1}{\alpha-1}} \tilde{L}(n)$.

In this case we have that the slowing is caused by the number of excursions in traps. Since β is small (i.e. less than μ^{-1}) we have that the expected time spent in a trap is finite. The number of excursions the walk takes into a branch is of the same order as the number of buds; since the size-biased law has infinite mean there are a large number of buds and therefore a large number of excursions. The main result for IVFE is Theorem 1 which reflects that Δ_n scales similarly to the sum of independent copies of ξ^*.

Theorem 1 *For IVFE, the laws of the process*

$$\left(\frac{\Delta_{nt}}{a_n}\right)_{t\geq 0}$$

converge weakly as $n \to \infty$ under $\mathbb{P}$ with respect to the Skorohod J_1 topology on $D([0, \infty), \mathbb{R})$ to the law of an $\alpha - 1$ stable subordinator R_t with Laplace transform

$$\varphi_t(s) := \mathbb{E}\left[e^{-sR_t}\right] = e^{-C_{\alpha,\beta,\mu} t s^{\alpha-1}}$$

where $C_{\alpha,\beta,\mu}$ is a constant which we shall determine during the proof (see 9.1).

We refer to the second ($\sigma^2 < \infty$, $\beta\mu > 1$) and third ($\sigma^2 = \infty$, $\beta\mu > 1$) cases as the finite variance, infinite excursion and infinite variance, infinite excursion cases respectively.

Definition 2 (*FVIE*) The offspring distribution has mean μ satisfying $1 < \mu^{-1} < \beta$ and variance $\sigma^2 < \infty$.

Definition 3 (*IVIE*) The offspring distribution has mean μ satisfying $1 < \mu^{-1} < \beta$ and belongs to the domain of attraction of a stable law of index $\alpha \in (1, 2)$.

As for IVFE, in IVIE we let L vary slowly at ∞ such that (2.2) holds and $(a_n)_{n\geq 1}$ be some scaling sequence for the size-biased law such that for any $x > 0$, as $n \to \infty$ we have $\mathbf{P}(\xi^* \geq x a_n) \sim x^{-(\alpha-1)} n^{-1}$. It then follows that $a_n = n^{\frac{1}{\alpha-1}} \tilde{L}(n)$ for some slowly varying function $\tilde{L}$. In FVIE, $a_n = n$ will suffice.

In FVIE and IVIE the slowing is caused by excursions in deep traps because the walk is required to make long sequences of movements against the bias in order to escape. We shall see that only the depth H (and not the foliage) is important to the scaling. By comparison with the model in which we strip all of the branch except the unique self-avoiding path to the deepest point; we see that, by transience, the walk reaches the deepest point with positive probability and then takes a geometric number of short excursions with escape probability of the order β^{-H}. In particular, this means that the expected time spent in a branch of height H will cluster around β^H.

Intuitively, the main reason we observe different scalings in these two cases is due to the way the number of buds affects the height of the branch. The height of a GW-tree is approximately geometric; in particular, the tallest of n independent trees will typically be close to $\log(n)/\log(\mu^{-1})$. In FVIE the number of buds has finite mean therefore we see order n buds by level n hence tallest will have height close to $\log(n)/\log(\mu^{-1})$. In IVIE the number of buds has infinite mean but belongs to the domain of attraction of some stable law. In particular, the number of buds seen by level n is equal in distribution to the sum of n independent copies of $\xi^* - 1$ (which scales with a_n). It therefore follows that, in IVIE, the tallest tree up to level n will have height close to $\log(a_n)/\log(\mu^{-1})$. Since only the deepest trees are significant and the time spent in a large branch clusters around β^H we see that the natural scaling is $\beta^{\log(n)/\log(\mu^{-1})} = n^{1/\gamma}$ in FVIE and $\beta^{\log(a_n)/\log(\mu^{-1})} = a_n^{1/\gamma}$ in IVIE.

Since H is approximately geometric we have that β^H won't belong to the domain of attraction of any stable law. For this reason, as in [4], we only see convergence along specific increasing subsequences $n_l(t) := \lfloor t\mu^{-l} \rfloor$ for $t > 0$ in FVIE and $n_l(t)$ such that $a_{n_l(t)} \sim t\mu^{-l}$ for IVIE. Such a sequence exists for any $t > 0$ since by choosing $n_l(t) := \sup\{m \geq 0 : a_m < t\mu^{-l}\}$ we have that $a_{n_l} < t\mu^{-l} \leq a_{n_l+1}$ and therefore

$$1 \geq \frac{a_{n_l}}{t\mu^{-l}} \geq \frac{a_{n_l}}{a_{n_l+1}} \to 1.$$

Recalling (1.1), the main results for FVIE and IVIE are Theorems 2 and 3, which reflect slowing due to deep excursions.

Theorem 2 *In FVIE, for any $t > 0$ we have that as $l \to \infty$*

$$\frac{\Delta_{n_l(t)}}{n_l(t)^{\frac{1}{\gamma}}} \to R_t$$

in distribution under $\mathbb{P}$, where R_t is a random variable with an infinitely divisible law.

Theorem 3 *In IVIE, for any $t > 0$ we have that as $l \to \infty$*

$$\frac{\Delta_{n_l(t)}}{a_{n_l(t)}^{\frac{1}{\gamma}}} \to R_t$$

in distribution under $\mathbb{P}$, where R_t is a random variable with an infinitely divisible law.

We write r_n to be a_n in IVFE, $n^{1/\gamma}$ in FVIE and $a_n^{1/\gamma}$ in IVIE; then, letting $\bar{r}_n := \max\{m \geq 0 : r_m \leq n\}$ we will also prove Theorem 4. This shows that, although the laws of $X_n/\bar{r}_n$ don't converge in general (for FVIE and IVIE), the suitably scaled sequence is tight and we can determine the leading order polynomial exponent explicitly.

Theorem 4 *In IVFE, FVIE or IVIE we have that*

1. *The laws of $(\Delta_n/r_n)_{n\geq 0}$ under $\mathbb{P}$ are tight on $(0, \infty)$;*
2. *The laws of $(|X_n|/\bar{r}_n)_{n\geq 0}$ under $\mathbb{P}$ are tight on $(0, \infty)$.*

Moreover, in IVFE, FVIE and IVIE respectively, we have that $\mathbb{P}$-a.s.

$$\lim_{n\to\infty} \frac{\log|X_n|}{\log(n)} = \alpha - 1;$$

$$\lim_{n\to\infty} \frac{\log|X_n|}{\log(n)} = \gamma;$$

$$\lim_{n\to\infty} \frac{\log|X_n|}{\log(n)} = \gamma(\alpha - 1).$$

The final case we consider is an extension of a result of [4] for the walk on the supercritical tree. The argument used for the infinite variance case is generally the same as in the finite variance case but needs some technical input. This is provided by three lemmas which we put aside until Sect. 10. For the same reason as in FVIE, we only see convergence along specific subsequences $n_l(t) := \lfloor tf'(q)^{-l} \rfloor$ for $t > 0$.

Theorem 5 (Infinite variance supercritical case) *Suppose the offspring law belongs to the domain of attraction of some stable law of index $\alpha \in (1, 2)$, has mean $\mu > 1$ and the bias satisfies the bound $\beta > f'(q)^{-1}$. Then,*

$$\frac{\Delta_{n_l(t)}}{n_l(t)^{\frac{1}{\gamma}}} \to R_t$$

in distribution as $l \to \infty$ under $\mathbb{P}$, where R_t is a random variable with an infinitely divisible law whose parameters are given in [4]. Moreover, the laws of $(\Delta_n n^{-\frac{1}{\gamma}})_{n\geq 0}$ and $(|X_n|n^{-\gamma})_{n\geq 0}$ under $\mathbb{P}$ are tight on $(0, \infty)$ and $\mathbb{P}$-a.s.

$$\lim_{n\to\infty} \frac{\log|X_n|}{\log(n)} = \gamma.$$

The proofs of Theorems 1, 2 and 3 follow a similar structure to the corresponding proof of [4] which, for the walk on the supercritical tree, only considers the case in which the variance of the offspring distribution is finite. However, for the latter reason, the proofs of Theorems 1 and 3 become more technical in some places, specifically with regards to the number of traps in a large branch. The proof can be broken down into a sequence of stages which investigate different aspects of the walk and the tree. This is ideal for extending the result onto the supercritical tree because many of these

behavioural properties will be very similar for the walk on the subcritical tree due to the similarity of the traps.

In all cases it will be important to decompose large branches. In Sect. 3 we show a decomposition of the number of deep traps in any deep branch. This is only important for FVIE and IVIE since the depth of the branch plays a key role in decomposing the time spent in large branches. In Sect. 4 we determine conditions for labelling a branch as large in each of the regimes so that large branches are sufficiently far apart so that, with high probability, the underlying walk won't backtrack from one large branch to the previous one. In Sect. 5 we justify the choice of label by showing that time spent outside these large branches is negligible. From this we then have that Δ_n can be approximated by a sum of i.i.d. random variables whose distribution depends on n. In Sect. 6 we only consider IVFE and show that, under a suitable scaling, these variables converge in distribution which allows us to show the convergence of their sum. Similarly, in Sect. 7 we show that the random variables, suitably scaled, converge in distribution for FVIE and IVIE. We then show convergence of their sum in Sect. 8. In Sect. 9 we prove Theorem 4 which is standard following Theorems 1, 2 and 3. Finally, in Sect. 10, we prove three short lemmas which extend the main result of [4] to prove Theorem 5. We require a lot of notation much of which is very similar; a glossary follows Sect. 10 which includes most of the notation used repeatedly throughout.

3 Number of traps

In this section we show asymptotics for the probability that the height of a branch is large and use it to determine the distribution over the number of large traps in a large branch. Unless stated otherwise we assume $\mu < 1$.

In the construction of the subcritical GW-tree conditioned to survive $\mathcal{T}$ described in the introduction, the special vertices form the infinite backbone $\mathcal{Y} = \{\rho_0, \rho_1, \ldots\}$ consisting of all vertices with an infinite line of descent where ρ_i is the vertex in generation i. Each vertex ρ_i on the backbone is connected to buds $\rho_{i,j}$ for $j = 1, \ldots, d_{\rho_i} - 1$ (which are the normal vertices that are offspring of special vertices in the construction). Each of these is then the root of an f-GW tree $\mathcal{T}_{\rho_{i,j}}$. We call each $\mathcal{T}_{\rho_{i,j}}$ a trap and the collection from a single backbone vertex (combined with the backbone vertex) $\mathcal{T}_{\rho_i}^{*-}$ a branch. Figure 2 shows an example of the first five generations of a tree $\mathcal{T}$. The solid line represents the backbone and the two dotted ellipses identify a sample branch and trap. The dashed ellipse indicates the children of ρ_1 which, since ρ_1 is on the backbone, have quantity distributed according to the size-biased law. It will be helpful throughout to work on a dummy branch which is equal in distribution to $\mathcal{T}_{\rho_i}^{*-}$ for any i thus we define the following random tree.

Definition 4 (*Dummy branch*) Define $\mathcal{T}^{*-}$ to be a random tree rooted at ρ with first generation vertices $\rho_1, \ldots, \rho_{\xi^*-1}$ which are roots of independent f-GW-trees $(\mathcal{T}_i^\circ)_{i=1}^k$ where ξ^* is a size biased random variable independent of the rest of the tree. Define $\mathcal{T}^\circ$ to be a dummy f-GW-tree.

The structure of the large traps will have an important role in determining the convergence of the scaled process. In this section we determine the distribution over
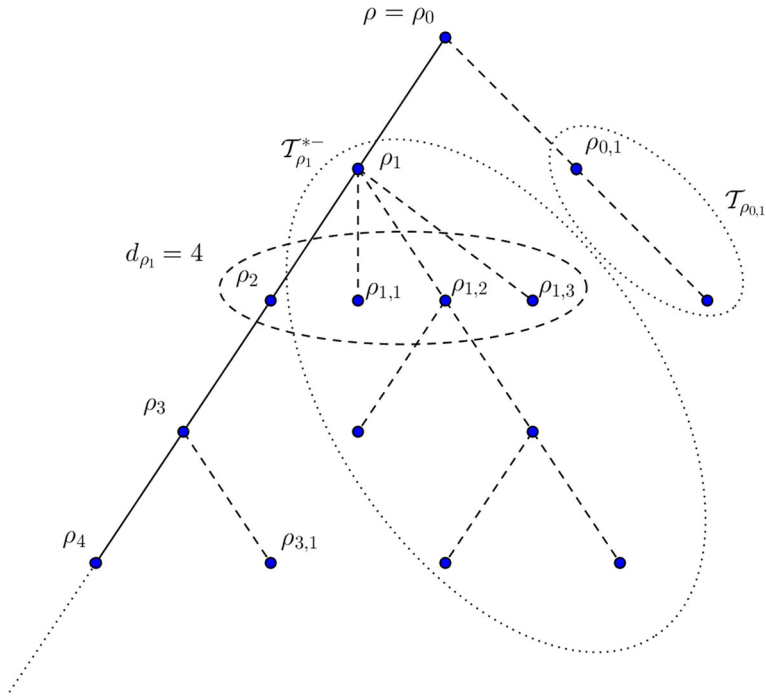
Fig. 2 A sample section of a subcritical tree conditioned to survive with *solid lines* representing the backbone $\mathcal{Y}$ and *dashed lines* representing the dangling ends

the number of deep traps rooted at backbone vertices with at least one deep trap. We will show that there is only a single deep trap at any backbone vertex when the offspring law has finite variance whereas, when the offspring law belongs to the domain of attraction of a stable law with index $\alpha < 2$ we have that the number of deep traps converges in distribution to a certain heavy tailed law.

A fundamental result for branching processes (see, for example [20]), is that for $\mu < 1$ and Z_n an f-GW process, the sequence $\mathbf{P}(Z_n > 0)/\mu^n$ is decreasing; moreover, $\mathbf{E}[\xi \log(\xi)] < \infty$ if and only if the limit of $\mathbf{P}(Z_n > 0)\mu^{-n}$ as $n \to \infty$ exists and is strictly positive. This assumption holds under any of the hypotheses thus for this paper we will always make this assumption and let c_μ be the constant such that

$$\mathbf{P}(Z_n > 0) \sim c_\mu \mu^n. \tag{3.1}$$

Recall that $\mathcal{H}(T)$ denotes the height of a tree T rooted at ρ. Denote

$$s_m := \mathbf{P}\left(\mathcal{H}(\mathcal{T}^\circ) < m\right) = 1 - c_\mu \mu^m (1 + o(1)) \tag{3.2}$$

to be the probability that a given trap is of height at most $m-1$ (although in general we shall write s for convenience). Write $N(m) := \sum_{j=1}^{\xi^*-1} \mathbf{1}_{\{\mathcal{H}(\mathcal{T}_i^\circ) \geq m\}}$ to be the number of traps of height at least m in the dummy branch then we are interested in the limit as $m \to \infty$ of

$$\mathbf{P}\left(N(m)=l\,\middle|\,N(m)\geq 1\right)=\frac{\mathbf{P}\left(N(m)=l\right)}{\mathbf{P}\left(N(m)\geq 1\right)} \tag{3.3}$$

for $l\geq 1$. Recall that f is the p.g.f. of the offspring distribution, write $f^{(k)}$ for its kth derivative then we have that

$$\mathbf{P}\left(N(m)=l\right)=\sum_{k=1}^{\infty}\mathbf{P}\left(\xi^*=k\right)\mathbf{P}\left(N(m)=l\,|\,\xi^*=k\right)$$

$$=\sum_{k=l+1}^{\infty}\frac{kp_k}{\mu}s^{k-1-l}(1-s)^l\binom{k-1}{l}$$

$$=\frac{(1-s)^l}{l!\mu}f^{(l+1)}(s). \tag{3.4}$$

In particular, we have that

$$\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-})>m\right)=\mathbf{P}\left(N(m)\geq 1\right)=1-f'(s)/\mu. \tag{3.5}$$

Lemma 3.1 shows that, when $\sigma^2<\infty$, with high probability there will only be a single deep trap in any deep branch.

Lemma 3.1 *When* $\sigma^2<\infty$

$$\lim_{m\to\infty}\mathbf{P}\left(N(m)=1\,\middle|\,N(m)\geq 1\right)=1.$$

Proof Using (3.3) and (3.4) we have that

$$\mathbf{P}\left(N(m)=1\,\middle|\,N(m)\geq 1\right)=\frac{(1-s)f''(s)/\mu}{1-f'(s)/\mu}=\frac{\sum_{k=2}^{\infty}k(k-1)p_k s^{k-2}}{\sum_{k=2}^{\infty}kp_k\frac{1-s^{k-1}}{1-s}}. \tag{3.6}$$

By monotonicity in s we have that

$$\lim_{s\to 1^\circ}\sum_{k=2}^{\infty}k(k-1)p_k s^{k-2}=\sum_{k=2}^{\infty}k(k-1)p_k$$

which is finite since $\sigma^2<\infty$. Each summand in the denominator is increasing in s for $s\in(0,1)$ and by L'Hopital's rule $1-s^{k-1}\sim(k-1)(1-s)$ as $s\to 1^\circ$ therefore, by monotone convergence, the denominator in the final term of (3.6) converges to the same limit. $\square$

In order to determine the correct threshold for labelling a branch as large we will need to know the asymptotic form of $\mathbf{P}(N(m)\geq 1)$. Corollary 3.2 gives this for the finite variance case.

Corollary 3.2 *Suppose $\sigma^2 < \infty$ then*

$$\mathbf{P}(N(m) \geq 1) \sim c_\mu \mathbf{E}\left[\xi^* - 1\right]\mu^m = c_\mu \left(\frac{\sigma^2 + \mu^2}{\mu} - 1\right)\mu^m.$$

Proof Let f_* denote the p.g.f. of ξ^* then $\mathbf{P}(N(m) \geq 1) = 1 - s^{-1}f_*(s)$. Since $\sigma^2 < \infty$ we have that $f'_*(s)$ exists and is continuous for $s \leq 1$ thus as $s \to 1^\circ$ we have that $f_*(1) - f_*(s) \sim (1-s)f'_*(1) = (1-s)\mathbf{E}[\xi^*]$. It therefore follows that

$$1 - s^{-1}f_*(s) = f_*(1) - f_*(s) - \frac{f_*(s)(1-s)}{s} \sim (1-s)\left(\mathbf{E}\left[\xi^*\right] - 1\right).$$

The result then follows by the definitions of c_μ (3.1) and s (3.2). $\qquad\qquad\square$

We now consider the case when $\sigma^2 = \infty$ but ξ belongs to the domain of attraction of a stable law of index $\alpha \in (1, 2)$. The following lemma concerning the form of the probability generating function of the offspring distribution will be fundamental in determining the distribution over the number of large traps rooted at a given backbone vertex. The case $\mu = 1$ appears in [7]; the proof of Lemma 3.3 is a simple extension of this hence the proof is omitted.

Lemma 3.3 *Suppose the offspring distribution belongs to the domain of attraction of a stable law with index $\alpha \in (1, 2)$ and mean $\mathbf{E}[\xi] = \mu$.*

1. *If $\mu \leq 1$ then as $s \to 1^\circ$*

$$\mathbf{E}\left[s^\xi\right] - s^\mu \sim \frac{\Gamma(3-\alpha)}{\alpha(\alpha-1)}(1-s)^\alpha L((1-s)^{-1})$$

   where $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}\mathrm{d}x$ is the usual gamma function.
2. *If $\mu > 1$ then*

$$1 - \mathbf{E}\left[s^\xi\right] = \mu(1-s) + \frac{\Gamma(3-\alpha)}{\alpha(\alpha-1)}(1-s)^\alpha \overline{L}\left((1-s)^{-1}\right)$$

   where $\overline{L}$ varies slowly at ∞.

When $\mu < 1$ it follows that there exists a function L_1 (which varies slowly as $s \to 1^\circ$) such that $\mathbf{E}[s^\xi] - s^\mu = (1-s)^\alpha L_1((1-s)^{-1})$ and

$$\lim_{s\to 1^\circ} \frac{L_1\left((1-s)^{-1}\right)}{L\left((1-s)^{-1}\right)} = \frac{\Gamma(3-\alpha)}{\alpha(\alpha-1)}. \tag{3.7}$$

Write $g(x) = x^\alpha L_1(x^{-1})$ so that $f(s) = s^\mu + g(1-s)$ then it follows that

$$f^{(l)}(s) = s^{\mu-l}(\mu)_l + (-1)^l g^{(l)}(1-s)$$

when this exists where $(\mu)_l := \prod_{j=0}^{l-1}(\mu - j)$ is the Pochhammer symbol. Write $L_2(x) := L_1(x^{-1})$ which is slowly varying at 0. Using Theorem 2 of [19], we see that $xg'(x) \sim \alpha g(x)$ as $x \to 0$. Moreover, using an inductive argument in the proof of this result, it is straightforward to show that for all $l \in \mathbb{N}$ we have that $xg^{(l+1)}(x) \sim (\alpha - l)g^{(l)}(x)$ as $x \to 0$. Therefore, for any integer $l \geq 0$

$$\lim_{x \to 0^+} \frac{x^l g^{(l)}(x)}{g(x)} = (\alpha)_l. \tag{3.8}$$

Proposition 3.4 is the main result of this section and determines the limiting distribution of the number of traps of height at least m in a branch of height greater than m.

Proposition 3.4 *In IVIE, for $l \geq 1$ as $m \to \infty$*

$$\mathbf{P}\left(N(m) = l | N(m) \geq 1\right) \to \frac{1}{l!} \prod_{j=1}^{l} |\alpha - j|.$$

Proof Recall that by (3.3) and (3.4) we want to determine the asymptotics of $1 - f'(s)/\mu$ and $(1 - s)^l f^{(l+1)}(s)/(l!\mu)$ as $s \to 1^\circ$. We have that $1 - f'(s)/\mu = 1 - s^{\mu-1} + g'(1 - s)/\mu$ and $g'(1 - s) \sim \alpha(1 - s)^{\alpha-1}L_2(1 - s)$ as $s \to 1$. Since $\alpha < 2$, we have that $\lim_{s \to 1^\circ}(1 - s^{\mu-1})(1 - s)^{1-\alpha} = 0$ hence

$$1 - \frac{f'(s)}{\mu} \sim \frac{\alpha}{\mu}(1 - s)^{\alpha-1}L_2(1 - s). \tag{3.9}$$

For derivatives $l \geq 1$ we have that

$$\frac{(1 - s)^l f^{(l+1)}(s)}{l!\mu} = \frac{(1 - s)^l}{l!\mu}\left(s^{\mu-(l+1)}(\mu)_l + (-1)^{l+1}g^{(l+1)}(1 - s)\right).$$

By (3.8) we have that $(1 - s)^l g^{(l+1)}(1 - s) \sim (\alpha)_{l+1}(1 - s)^{\alpha-1}L_2(1 - s)$. For $l \geq 1$ we have that $l + 1 - \alpha > 0$ hence

$$\frac{(1 - s)^l f^{(l+1)}(s)}{l!\mu} \sim \frac{|(\alpha)_{l+1}|}{l!\mu}(1 - s)^{\alpha-1}L_2(1 - s). \tag{3.10}$$

Combining (3.3) with (3.9) and (3.10) gives the desired result. □

Proposition 3.4 will be useful for determining the number of large traps in a large branch but equally important is the asymptotic relation (3.9) which gives the tail behaviour of the height of a branch $\mathcal{T}^{*-}$. By the assumption on ξ that (2.2) holds we have that

$$\mathbf{P}\left(\xi^* \geq t\right) \sim \frac{2 - \alpha}{\mu(\alpha - 1)}t^{-(\alpha-1)}L(t) \tag{3.11}$$

as $t \to \infty$. Using (3.2), (3.5), (3.7), (3.9) and (3.11), we then have that

$$\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) > m\right) \sim \frac{\Gamma(3-\alpha)c_\mu^{\alpha-1}}{\mu(\alpha-1)} \mu^{m(\alpha-1)} L(\mu^{-m}) \sim \Gamma(2-\alpha)c_\mu^{\alpha-1} \mathbf{P}\left(\xi^* \geq \mu^{-m}\right).$$
(3.12)

4 Large branches are far apart

In this section we introduce the conditions for a branch to be large. This will differ in each of the cases however, since many of the proofs will generalise to all three cases, we will use the same notation for some aspects.

In IVFE we will have that the slowing is caused by the large number of traps. In particular, we will be able to show that the time spent outside branches with a large number of buds is negligible.

Definition 5 (*IVFE large branch*) For $\varepsilon \in (0, 1)$ write

$$l_{n,\varepsilon} := a_{\lfloor n^{1-\varepsilon} \rfloor} \quad \text{and} \quad l_{n,\varepsilon}^+ := a_{\lfloor n^{1+\varepsilon} \rfloor}$$

then we have that $\mathbf{P}(\xi^* \geq l_{n,\varepsilon}) \sim n^{-(1-\varepsilon)}$. We will call a branch large if the number of buds is at least $l_{n,\varepsilon}$ and write $\mathcal{D}^{(n)} := \{x \in \mathcal{Y} : d_x > l_{n,\varepsilon}\}$ to be the collection of backbone vertices which are the roots of large branches.

In FVIE we will have that the slowing is caused by excursions into deep traps.

Definition 6 (*FVIE large branch*) For $\varepsilon \in (0, 1)$ write $C_\mathcal{D} := c_\mu \mathbf{E}[\xi^* - 1]$,

$$h_{n,\varepsilon} := \left\lceil \frac{(1-\varepsilon)\log(n)}{\log(\mu^{-1})} \right\rceil \quad \text{and} \quad h_{n,\varepsilon}^+ := \left\lceil \frac{(1+\varepsilon)\log(n)}{\log(\mu^{-1})} \right\rceil$$

then by Corollary 3.2 we have that

$$\mathbf{P}\left(\mathcal{H}\left(\mathcal{T}^{*-}\right) > h_{n,\varepsilon}\right) \sim C_\mathcal{D} \mu^{h_{n,\varepsilon}} \approx C_\mathcal{D} n^{-(1-\varepsilon)}.$$

We will call a branch large if there exists a trap within it of height at least $h_{n,\varepsilon}$ and write $\mathcal{D}^{(n)} := \{x \in \mathcal{Y} : \mathcal{H}(\mathcal{T}_x^{*-}) > h_{n,\varepsilon}\}$ to be the collection of backbone vertices which are the roots of large branches. By a large trap we mean any trap of height at least $h_{n,\varepsilon}$.

In IVIE we will have that the slowing is caused by a combination of the slowing effects of the other two cases. The height and number of buds in branches have a strong link which we show more precisely later; this allows us to label branches as large based on height which will be necessary when decomposing the time spent in large branches.

Definition 7 (*IVIE large branch*) For $\varepsilon \in (0, 1)$ write

$$h_{n,\varepsilon} := \left\lceil \frac{\log\left(a_{n^{1-\varepsilon}}\right)}{\log\left(\mu^{-1}\right)} \right\rceil \quad \text{and} \quad h_{n,\varepsilon}^+ := \left\lceil \frac{\log\left(a_{n^{1+\varepsilon}}\right)}{\log\left(\mu^{-1}\right)} \right\rceil$$

then by (3.12), for $C_{\mathcal{D}} := \Gamma(2 - \alpha)c_{\mu}^{\alpha-1}$, we have that

$$\mathbf{P}\left(\mathcal{H}\left(\mathcal{T}^{*-}\right) > h_{n,\varepsilon}\right) \sim C_{\mathcal{D}}\mathbf{P}\left(\xi^* \geq \mu^{-h_{n,\varepsilon}}\right) \approx C_{\mathcal{D}}n^{-(1-\varepsilon)}. \tag{4.1}$$

We will call a branch large if there exists a trap of height at least $h_{n,\varepsilon}$ and write $\mathcal{D}^{(n)} := \{x \in \mathcal{Y} : \mathcal{H}(\mathcal{T}_x^{*-}) > h_{n,\varepsilon}\}$ to be the collection of backbone vertices which are the roots of large branches. By a large trap we mean any trap of height at least $h_{n,\varepsilon}$.

We want to show that, asymptotically, the large branches are sufficiently far apart to ignore any correlation and therefore approximate Δ_n by the sum of i.i.d. random variables representing the time spent in a large branch. Much of this is very similar to [4] so we only give brief details.

Write $\mathcal{D}_m^{(n)} := \{x \in \mathcal{D}^{(n)} : |x| \leq m\}$ to be the large roots before level m then let $q_n := \mathbf{P}(\rho \in \mathcal{D}^{(n)})$ be the probability that a branch is large and write

$$A_1(n, T) := \left\{ \sup_{t\in[0,T]} \left| \left|\mathcal{D}_{\lfloor tn \rfloor}^{(n)}\right| - \lfloor tnq_n \rfloor \right| < n^{2\varepsilon/3} \right\} \tag{4.2}$$

to be the event that the number of large branches by level Tn doesn't differ too much from its expected value. Notice that in all three cases we have that q_n is of the order $n^{-(1-\varepsilon)}$ thus we expect to see $nq_n \approx Cn^{\varepsilon}$ large branches by level n.

Lemma 4.1 *For any $T > 0$*

$$\lim_{n\to\infty} \mathbf{P}\left(A_1(n, T)^c\right) = 0.$$

Proof For each $n \in \mathbb{N}$ write

$$M_m^n := \left|\mathcal{D}_m^{(n)}\right| - mq_n \overset{\mathrm{d}}{=} \sum_{k=1}^{m} (B_k - q_n)$$

where B_k are independent Bernoulli random variables with success probability q_n. Then $\mathbf{E}[M_m^n] = 0$ and $Var_{\mathbf{P}}(M_m^n) = mq_n(1 - q_n)$ therefore by Kolmogorov's maximal inequality

$$\mathbf{P}\left(\max_{1\leq m\leq\lfloor nT \rfloor} |M_m^n| \geq n^{2\varepsilon/3} - 1\right) \leq \frac{cVar_{\mathbf{P}}\left(M_{\lfloor nT \rfloor}^n\right)}{n^{4\varepsilon/3}} \leq \frac{cnTq_n}{n^{4\varepsilon/3}} \leq CTn^{-\varepsilon/3}.$$

Since $||\lfloor nt \rfloor q_n - \lfloor ntq_n \rfloor|| \leq 1$ we have that

$$
\begin{aligned}
\sup_{t\in[0,T]} \left|\left|\mathcal{D}^{(n)}_{\lfloor tn \rfloor}\right| - \lfloor tnq_n \rfloor\right| &\leq \max_{1\leq m\leq \lfloor nT \rfloor} \left|M^n_m\right| + \sup_{0\leq t\leq T} |tnq_n - \lfloor nt \rfloor q_n| \\
&\leq \max_{1\leq m\leq \lfloor nT \rfloor} \left|M^n_m\right| + 1
\end{aligned}
$$

which proves the statement. $\qquad\square$

We want to show that all of the large branches are sufficiently far apart such that the walk doesn't backtrack from one to another. For $t > 0$ and $\kappa \in (0, 1 - 2\varepsilon)$ write

$$
\mathcal{D}(n, t) := \left\{ \min_{x,y\in\mathcal{D}^{(n)}_{\lfloor nt \rfloor} : x\neq y} d(x, y) > n^{\kappa} \right\} \cap \left\{ \rho \notin \mathcal{D}^{(n)} \right\}
$$

to be the event that all large branches up to level $\lfloor nt \rfloor$ are of distance at least n^{κ} apart and the root of the tree is not the root of a large branch. A union bound shows that $\mathbf{P}(\mathcal{D}(n, t)^c) \to 0$ as $n \to \infty$ uniformly over t in compact sets.

We want to show that, with high probability, once the walk reaches a large branch it never backtracks to the previous one. For $t > 0$ write

$$
A^{(0)}_2(n, t) := \bigcap_{i=0}^{\lfloor nt \rfloor} \bigcap_{m\geq \Delta^Y_{\rho_i}} \left\{ |Y_m| > i - \overline{C}\log(n) \right\}
$$

to be the event that the walk never backtracks distance $\overline{C}\log(n)$ (where $\Delta^Y_n := \min\{m \geq 0 : Y_m = \rho_n\}$). For $x \in \mathcal{T}$ write $\tau^+_x = \inf\{n > 0 : X_n = x\}$ to be the first return time of x. Comparison with a simple random walk on $\mathbb{Z}$ shows that for $k \geq 1$ we have that the escape probability is $P_{\rho_k}\left(\tau^+_{\rho_{k-1}} < \infty\right) = \beta^{-1}$ hence, using the Strong Markov property,

$$
P_{\rho_m}\left(\tau^+_{\rho_0} < \infty\right) = \beta^{-m}.
$$

Using a union bound we see that

$$
\mathbb{P}\left(A^{(0)}_2(n, t)^c\right) \leq Cnt\beta^{-\overline{C}\log(n)} \to 0 \tag{4.3}
$$

for $\overline{C}$ sufficiently large. Combining this with $\mathcal{D}(n, t)$ we have that with high probability the walk never backtracks from one large branch to a previous one.

5 Time is spent in large branches

In this section we show that the time spent up to time Δ_n outside large branches is negligible. Combined with Sect. 4 this allows us to approximate Δ_n by the sum of

i.i.d. random variables. We begin with some general results concerning the number of excursions into traps and the expected time spent in a trap of height at most m.

Recall that $\rho_{i,j}$ are the buds connected to the backbone vertex ρ_i, that $c(\rho_i) := \{\rho_{i,j}\}_j \cup \{\rho_{i+1}\}$ is the collection of all offspring of ρ_i and $d_{\rho_i} = |c(\rho_i)|$ is the number of offspring. We write $W^{i,j} := |\{m \geq 0 : X_{m-1} = \rho_i, \ X_m = \rho_{i,j}\}|$ to be the number of excursions into the jth trap of the ith branch where we set $W^{i,j} := 0$ if $\rho_{i,j}$ doesn't exist in the tree. Lemma 5.1 shows that, conditional on the number of buds, the number of excursions follows a geometric law.

Lemma 5.1 *For any $i, k \in \mathbb{N}$ and $A \subset \{1, \ldots, d_{\rho_i} - 1\}$, when $\beta > 1$*

$$\sum_{j \in A} W^{i,j} \sim Geo\left(\frac{\beta - 1}{(|A| + 1)\beta - 1}\right)$$

with respect to $P^{\mathcal{T}}$. In particular for any $j \leq k$ we have that $W^{i,j} \sim Geo(p)$ where $p = (\beta - 1)/(2\beta - 1)$.

Moreover, under this law, $(W^{i,j})_{j \in A}$ have a negative multinomial distribution with one failure until termination and probabilities

$$q_j := \begin{cases} \frac{\beta - 1}{(|A| + 1)\beta - 1} & j = 0 \\[2mm] \frac{\beta}{(|A| + 1)\beta - 1} & j \in A \end{cases}$$

that from ρ_i the next excursion will be into the jth trap (where $j = 0$ denotes escaping).

Proof From $\rho_{i,j}$ the walk must return to ρ_i before escaping therefore since $P_{\rho_{i,j}}(\tau_{\rho_i}^+ < \infty) = 1$, any traps not in the set we consider can be ignored and it suffices to assume that $A = \{1, \ldots, k\}$. By comparison with a biased random walk on $\mathbb{Z}$ we have that $P_{\rho_{i+1}}(\tau_{\rho_i}^+ = \infty) = 1 - \beta^{-1}$. If $d_{\rho_i} = k+1$ then $P_{\rho_i}(\tau_x^+ = \min_{y \in c(\rho_i)} \tau_y^+) = (k+1)^{-1}$ for any $x \in c(\rho_i)$. The probability of never entering a trap in the branch $\mathcal{T}_{\rho_i}^{*-}$ is, therefore,

$$P_{\rho_i}\left(\bigcap_{j=1}^{k}\{\tau_{\rho_{i,j}}^+ = \infty\}\right) = \sum_{l=0}^{\infty}\left(\frac{1}{k+1}\beta^{-1}\right)^l\left(\frac{1 - \beta^{-1}}{k+1}\right) = \frac{\beta - 1}{(k+1)\beta - 1}.$$

Each excursion ends with the walker at ρ_i thus the walk takes a geometric number of excursions into traps with escape probability $(\beta - 1)/((k+1)\beta - 1)$. The second statement then follows from the fact that the walker has equal probability of going into any of the traps. □

For a fixed tree T with nth generation size Z_n^T where $Z_1^T > 0$ it is classical (e.g. [22]) that

$$E_\rho^T\left[\tau_\rho^+\right] = 2\sum_{n \geq 1}\frac{Z_n^T \beta^{n-1}}{Z_1^T}. \tag{5.1}$$

Let $\mathcal{T}^{\leftarrow}$ be an f-GW tree rooted at $\overleftarrow{\rho}$ conditioned to have a single first generation vertex which we label ρ. Notice that this has the same distribution as an f-GW-tree $\mathcal{T}^{\circ}$ to which we append a single ancestor of the root. From (5.1) it follows that

$$E_\rho^{\mathcal{T}^{\leftarrow}}\left[\tau_{\overleftarrow{\rho}}^+\right] = E_{\overleftarrow{\rho}}^{\mathcal{T}^{\leftarrow}}\left[\tau_{\overleftarrow{\rho}}^+\right] - 1 = 2\sum_{n\geq 0} Z_n^{\mathcal{T}^{\circ}}\beta^n - 1.$$

For any $m \geq 1$ we have that $\mathbf{P}(\mathcal{H}(\mathcal{T}^{\leftarrow}) \leq m) \geq p_0$ therefore, for some constant C,

$$\mathbf{E}\left[E_\rho^{\mathcal{T}^{\leftarrow}}\left[\tau_{\overleftarrow{\rho}}^+\right]\Big|\mathcal{H}(\mathcal{T}^{\leftarrow})\leq m\right] \leq \frac{\mathbf{E}\left[2\sum_{n=0}^{m-1} Z_n^{\mathcal{T}^{\circ}}\beta^n - 1\right]}{\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{\leftarrow})\leq m\right)} \leq \begin{cases} C(\mu\beta)^m & \beta\mu > 1 \\ Cm & \beta\mu = 1 \\ C & \beta\mu < 1. \end{cases}$$
$$(5.2)$$

Recall that Δ_n^Y is the first hitting time of ρ_n for the underlying walk Y and write

$$A_3(n) := \left\{\Delta_n^Y \leq C_1 n\right\}$$

to be the event that level n is reached by time $C_1 n$ by the walk on the backbone. Standard large deviation estimates yield that $\lim_{n\to\infty}\mathbb{P}(A_3(n)^c) = 0$ for $C_1 > (\beta+1)/(\beta-1)$.

For the remainder of this section we mainly consider the case in which ξ belongs to the domain of attraction of a stable law of index $\alpha \in (1, 2)$. The case in which the offspring law has finite variance will proceed similarly however since the corresponding estimates are much simpler in this case we omit the proofs.

In IVIE and IVFE, for $t > 0$, let the event that there are at most $\log(n)a_n$ buds by level $\lfloor nt \rfloor$ be

$$A_4(n, t) := \left\{\sum_{k=0}^{\lfloor nt \rfloor}(d_{\rho_k} - 1) \leq \log(n)a_n\right\}.$$

The variables d_{ρ_k} are i.i.d. with the law of ξ^* therefore the laws of $a_{\lfloor nt \rfloor}^{-1}\sum_{k=1}^{\lfloor nt \rfloor}(\xi_k^* - 1)$ converge to some stable law G^* where $\lim_{n\to\infty}\overline{G}^*(Ct^{\alpha-1}\log(n)) = 0$ uniformly over $t \leq T$ therefore we have that $\lim_{n\to\infty}\mathbf{P}(A_4(n, t)^c) = 0$.

In FVIE write

$$A_4(n, t) := \left\{\sum_{k=0}^{\lfloor nt \rfloor}(d_{\rho_k} - 1) \leq \log(n)n\right\}$$

then Markov's inequality gives that $\lim_{n\to\infty}\mathbf{P}(A_4(n, t)^c) = 0$.

Write

$$A_5(n) := \left\{\max_{i,j}\left|\{k \leq \Delta_{\lfloor nt \rfloor} : X_{k-1} = \rho_i, \ X_k = \rho_{i,j}\}\right| \leq C_2\log(n)\right\} \qquad (5.3)$$

to be the event that any trap is entered at most $C_2 \log(n)$ times. By Lemma 5.1 the number of entrances into $\rho_{i,j}$ has the law of a geometric random variable of parameter $p = (\beta - 1)/(2\beta - 1)$ hence using a union bound we have that

$$\mathbb{P}\left(A_5(n,t)^c \cap A_4(n,t)\right) \leq \log(n) a_n \mathbb{P}\left(Geo(p) > C_2 \log(n)\right) \leq \overline{L}(n) n^{\frac{1}{\alpha-1} + C_2 \log(1-p)}$$

where $\overline{L}$ varies slowly hence the final term converges to 0 for C_2 large and $\lim_{n \to \infty} \mathbb{P}(A_5(n,t)^c) = 0$.

Propositions 5.2, 5.3 and 5.4 show that any time spent outside large traps is negligible. In FVIE and IVIE we only consider the large traps in large branches. Recall that $\mathcal{D}^{(n)}$ is the set of roots of large branches and write

$$K(n) := \bigcup_{x \in \mathcal{D}^{(n)}} \left\{\mathcal{T}_y : y \in c(x) \backslash \left\{\rho_{|x|+1}\right\}, \mathcal{H}(\mathcal{T}_y) \geq h_{n,\varepsilon}\right\}$$

to be the vertices in large traps. In IVFE we require the entire large branch and write

$$K(n) := \bigcup_{x \in \mathcal{D}^{(n)}} \{y \in \mathcal{T}_x^{*-}\}$$

to be the vertices in large branches. In either case we write $\chi_{t,n} := |\{1 \leq i \leq \Delta_{\lfloor nt \rfloor} : X_{i-1}, X_i \in K(n)\}|$ to be the time spent up to $\Delta_{\lfloor nt \rfloor}$ in large traps.

Proposition 5.2 *In IVIE, for any $t, \epsilon > 0$ we have that as $n \to \infty$*

$$\mathbb{P}\left(\left|\frac{\Delta_{\lfloor nt \rfloor} - \chi_{t,n}}{a_n^{1/\gamma}}\right| \geq \epsilon\right) \to 0.$$

Proof On $A_4(n,t)$ there are at most $a_n \log(n)$ traps by level $\lfloor nt \rfloor$. We can order these traps so write $T^{(l,k)}$ to be the duration of the kth excursion into the lth trap and $\rho(l)$ to be the root of this trap (that is, the unique bud in the trap). Here we consider an excursion to start from the bud and end at the last hitting time of the bud before returning to the backbone. Recall that on $A_3(n)$ the walk Y reaches level n by time $C_1 n$ and on $A_5(n)$ no trap up to level n is entered more than $C_2 \log(n)$ times. Using the estimates on A_3, A_4 and A_5 we have that

$$\mathbb{P}\left(\left|\frac{\Delta_{\lfloor nt \rfloor} - \chi_{t,n}}{a_n^{1/\gamma}}\right| \geq \epsilon\right) \leq o(1)$$

$$+ \mathbb{P}\left(C_1 nt + \sum_{l=0}^{a_n \log(n)} \sum_{k=1}^{C_2 \log(nt)} \left(2 + T^{(l,k)} \mathbf{1}_{\{\mathcal{H}(\mathcal{T}_{\rho(l)}) < h_{n,\varepsilon}\}}\right) \geq \epsilon a_n^{\frac{1}{\gamma}}\right).$$

Since $a_n^{\frac{1}{\gamma}} \gg a_n \log(n)^2 \gg n$, for n sufficiently large we have that, using Markov's inequality and (5.2) with $m = h_{n,\varepsilon}$, the second term can be bounded above by

$$2\epsilon^{-1} a_n^{-\frac{1}{\gamma}} \mathbb{E}\left[\sum_{l=0}^{a_n \log(n)} \sum_{k=1}^{C_2 \log(nt)} T^{(l,k)} \mathbf{1}_{\{\mathcal{H}(\mathcal{T}_{\rho(l)}) < h_{n,\varepsilon}\}} \right] \leq C_{t,\epsilon} \log(n)^2 a_n^{1-\frac{1}{\gamma}} a_{n^{1-\varepsilon}}^{\frac{1}{\gamma}-1}.$$

Combining constants and slowly varying functions into a single function $L_{t,\epsilon}$ such that for any $\tilde{\varepsilon} > 0$ we have that $L_{t,\epsilon}(n) \leq n^{\tilde{\varepsilon}}$ for n sufficiently large we then have that

$$\mathbb{P}\left(\left| \frac{\Delta_{\lfloor nt \rfloor} - \chi_{t,n}}{a_n^{1/\gamma}} \right| \geq \epsilon \right) \leq o(1) + L_{t,\epsilon}(n) n^{-\varepsilon \frac{\frac{1}{\gamma}-1}{\alpha-1}}$$

which converges to 0 since $\alpha, \frac{1}{\gamma} > 1$.   □

Using A_3, A_5 and the form of A_4 for FVIE, the technique used to prove Proposition 5.2 extends straightforwardly to prove Proposition 5.3 therefore we omit the proof.

Proposition 5.3 *In FVIE, for any $t, \epsilon > 0$ we have that as $n \to \infty$*

$$\mathbb{P}\left(\left| \frac{\Delta_{\lfloor nt \rfloor} - \chi_{t,n}}{n^{1/\gamma}} \right| \geq \epsilon \right) \to 0.$$

Similarly, we can show a corresponding result for IVFE.

Proposition 5.4 *In IVFE, for any $t, \epsilon > 0$, as $n \to \infty$*

$$\mathbb{P}\left(\left| \frac{\Delta_{\lfloor nt \rfloor} - \chi_{t,n}}{a_n} \right| \geq \epsilon \right) \to 0.$$

Proof We begin by bounding the total number of traps in small branches. Recall from Definition 5 that $l_{n,\varepsilon} \leq a_{n^{1-\varepsilon}}$. Let $c \in (0, 2-\alpha)$ then, by Markov's inequality and the truncated first moment asymptotic:

$$\mathbf{E}\left[\xi^* \mathbf{1}_{\{\xi^* \leq x\}} \right] \sim C x^{2-\alpha} L(x) \tag{5.4}$$

as $x \to \infty$ for some constant C (see for example [12], IX.8), for n large

$$\mathbb{P}\left(\sum_{k=0}^{\lfloor nt \rfloor} (d_{\rho_k} - 1) \mathbf{1}_{\{d_{\rho_k} - 1 \leq l_{n,\varepsilon}\}} \geq n^{\frac{1-c\varepsilon}{\alpha-1}} \right) \leq \frac{\mathbf{E}\left[\sum_{k=0}^{\lfloor nt \rfloor} (d_{\rho_k} - 1) \mathbf{1}_{\{d_{\rho_k} - 1 \leq l_{n,\varepsilon}\}} \right]}{n^{\frac{1-c\varepsilon}{\alpha-1}}}$$

$$\leq n^{-\frac{\varepsilon(2-\alpha+c)}{\alpha-1}} L_t(n)$$

where $L_t(n)$ varies slowly at ∞. This converges to 0 as $n \to \infty$. We can order the traps in small branches and write $T^{(l,k)}$ to be the duration of the kth excursion in the

lth trap not in a large branch where we consider an excursion to start and end at the backbone. Using A_3 and A_5 to bound the time taken by Y to reach level nt and the number of entrances into traps up to level nt we have that for n suitably large

$$\mathbb{P}\left(\left|\frac{\Delta_{\lfloor nt \rfloor} - \chi_{t,n}}{a_n}\right| \geq \epsilon\right) \leq o(1) + \mathbb{P}\left(\sum_{l=0}^{n^{\frac{1-c\epsilon}{\alpha-1}} C_2 \log(nt)} \sum_{k=0}^{T^{(l,k)}} \geq \frac{\epsilon}{2} a_n\right).$$

Using Markov's inequality on the final term yields

$$\mathbb{P}\left(\sum_{l=0}^{n^{\frac{1-c\epsilon}{\alpha-1}} C_2 \log(nt)} \sum_{k=0} T^{(l,k)} \geq \frac{\epsilon}{2} a_n\right) \leq 2\epsilon^{-1} a_n^{-1} \mathbb{E}\left[\sum_{k=0}^{n^{\frac{1-c\epsilon}{\alpha-1}} C_2 \log(nt)} \sum_{j=0} T^{(l,k)}\right]$$

$$\leq n^{\frac{-c\epsilon}{\alpha-1}} L_{t,\epsilon}(n)$$

for some $L_{t,\epsilon}$ varying slowly at ∞. This converges to 0 as $n \to \infty$ hence the result holds. $\qquad\square$

Recall that we write r_n to be a_n in IVFE, $n^{1/\gamma}$ in FVIE and $a_n^{1/\gamma}$ in IVIE. Since $\Delta_{\lfloor nt \rfloor} - \chi_{t,n}$ is non-negative and non-decreasing in t we have that $\sup_{0 \leq t \leq T} |\Delta_{\lfloor nt \rfloor} - \chi_{t,n}| = |\Delta_{\lfloor nT \rfloor} - \chi_{T,n}|$ therefore Corollary 5.5 follows from Propositions 5.2, 5.3 and 5.4.

Corollary 5.5 *In each of IVFE, FVIE and IVIE, for any $T > 0$*

$$\sup_{0 \leq t \leq T} \frac{\left|\Delta_{\lfloor nt \rfloor} - \chi_{t,n}\right|}{r_n}$$

converges in $\mathbb{P}$-probability to 0.

Let Λ be the set of strictly increasing continuous functions mapping $[0, T]$ onto itself and I the identity map on $[0, T]$ then we consider the Skorohod J_1 metric

$$d_{J_1}(f, g) := \inf_{\lambda \in \Lambda} \sup_{t \in [0,T]} \left(|f(t) - g(\lambda(t))| + |t - \lambda(t)|\right).$$

Write χ_n^i to be the total time spent in large traps of the ith large branch; that is

$$\chi_n^i := \left|\left\{m \geq 0 : X_{m-1}, X_m \in \left(\mathcal{T}_{\rho_i^+}^{*-} \cap K(n)\right)\right\}\right|$$

where ρ_i^+ is the element of $\mathcal{D}^{(n)}$ which is ith closest to ρ. Notice that, whereas $\chi_{n,t}$ only accumulates time up to reaching $\rho_{\lfloor nt \rfloor}$, each χ_n^i may have contributions at arbitrarily large times. Recall that $A_2^{(0)}(n, t)$ is the event that the walk never backtracks distance

$\overline{C} \log(n)$ along the backbone from a backbone vertex up to level $\lfloor nt \rfloor$. On $A_2^{(0)}(n, T)$ we therefore have that for all $t \leq T$

$$\sum_{i=1}^{\left|\mathcal{D}_{\lfloor nt - \overline{C} \log(n) \rfloor}^{(n)}\right|} \chi_n^i \leq \chi_{n,t} \leq \sum_{i=1}^{\left|\mathcal{D}_{\lfloor nt \rfloor}^{(n)}\right|} \chi_n^i$$

where, on $\mathcal{D}(n, t)$, the J_1 distance between the two sums in the above expression can be bounded above by $\overline{C} \log(n)$. In particular, using that $A_2^{(0)}(n, T)$ and $\mathcal{D}(n, t)$ occur with high probability with the tightness result we prove in Sect. 9, in order to prove Theorems 1, 2 and 3 it will suffice to consider the time spent in large traps up to level $\lfloor nt \rfloor$ under the appropriate scaling.

Let $(X_n^{(i)})_{i \geq 1}$ be independent walks on $\mathcal{T}$ with the law of X_n and $(Y_n^{(i)})_{i \geq 1}$ the corresponding backbone walks. For $i \geq 1$ let $\tilde{\chi}_n^i$ be the time spent in the ith large trap by $X_n^{(i)}$ and

$$\tilde{\chi}_{t,n} := \sum_{i=1}^{\lfloor ntq_n \rfloor} \tilde{\chi}_n^i.$$

The random variables $\tilde{\chi}_n^i$ are independent copies (under $\mathbb{P}$) of times spent in large branches; moreover, on $\mathcal{D}(n, t)$, $\rho \notin \mathcal{D}^{(n)}$ therefore they are identically distributed. Let $\mathbb{E}$ extend to the enlarged space.

Lemma 5.6 *In each of IVFE, FVIE and IVIE,*

1. *as $n \to \infty$*

$$d_{J_1}\left(\left(\sum_{i=1}^{\left|\mathcal{D}_{\lfloor nt \rfloor}^{(n)}\right|} \frac{\chi_n^i}{r_n}\right)_{t \in [0,T]}, \left(\sum_{i=1}^{\lfloor tnq_n \rfloor} \frac{\chi_n^i}{r_n}\right)_{t \in [0,T]}\right)$$

   converges to 0 in probability;
2. *for any bounded $H : D([0, T], \mathbb{R}) \to \mathbb{R}$ continuous with respect to the Skorohod J_1 topology, as $n \to \infty$*

$$\left|\mathbb{E}\left[H\left(\left(\sum_{i=1}^{\lfloor tnq_n \rfloor} \frac{\chi_n^i}{r_n}\right)_{t \in [0,T]}\right)\right] - \mathbb{E}\left[H\left(\left(\sum_{i=1}^{\lfloor tnq_n \rfloor} \frac{\tilde{\chi}_n^i}{r_n}\right)_{t \in [0,T]}\right)\right]\right| \to 0.$$

Proof By definition of d_{J_1}, the distance in statement 1 is equal to

$$\inf_{\lambda \in \Lambda} \sup_{t \in [0,T]}\left(\left|\sum_{i=1}^{\left|\mathcal{D}_{\lfloor nt \rfloor}^{(n)}\right|} \frac{\chi_n^i}{r_n} - \sum_{i=1}^{\lfloor \lambda(t)nq_n \rfloor} \frac{\chi_n^i}{r_n}\right| + |\lambda(t) - t|\right).$$

For $m \in \mathbb{N}$ let $\lambda_n(m/n) := |\mathcal{D}_m^{(n)}|(nq_n)^{-1}$ then define $\lambda_n(t)$ by the usual linear interpolation. It follows that $|\mathcal{D}_{\lfloor nt \rfloor}^{(n)}| = \lfloor \lambda_n(t)nq_n \rfloor$ and the above expression can be bounded above by

$$
\sup_{t \in [0,T]} \left| t - \frac{\left| \mathcal{D}_{\lfloor nt \rfloor}^{(n)} \right|}{nq_n} \right|
$$

which converges to 0 by Lemma 4.1 since $n^{2\varepsilon/3}(nq_n)^{-1} \to 0$.

For $i \geq 1$ let

$$
A_2^{(i)}(n, t) := \bigcap_{j=0}^{\lfloor nt \rfloor} \bigcap_{m \geq \Delta_{\rho j}^{Y^{(i)}}} \left\{ \left| Y_m^{(i)} \right| > j - \overline{C} \log(n) \right\}
$$

be the analogue of $A_2^{(0)}(n, t)$ for the ith copy and

$$
\tilde{A}_2(n, t) := \mathcal{D}(n, t) \cap \bigcap_{i=0}^{\lfloor ntq_n \rfloor} A_2^{(i)}(n, t)
$$

be the event that ρ is not the root of a large branch, on each of the first $\lceil ntq_n \rceil$ copies the walk never backtracks distance $\overline{C} \log(n)$ and that large branches are of distance at least n^κ apart.

$$
\mathbb{E}\left[H\left(\left(\sum_{i=1}^{\lfloor tnq_n \rfloor} \frac{\chi_n^i}{r_n} \right)_{t \in [0,T]} \right) \mathbf{1}_{\tilde{A}_2(n,T)} \right] = \mathbb{E}\left[H\left(\left(\sum_{i=1}^{\lfloor tnq_n \rfloor} \frac{\tilde{\chi}_n^i}{r_n} \right)_{t \in [0,T]} \right) \mathbf{1}_{\tilde{A}_2(n,T)} \right]
$$

therefore

$$
\left| \mathbb{E}\left[H\left(\left(\sum_{i=1}^{\lfloor tnq_n \rfloor} \frac{\chi_n^i}{r_n} \right)_{t \in [0,T]} \right) \right] - \mathbb{E}\left[H\left(\left(\sum_{i=1}^{\lfloor tnq_n \rfloor} \frac{\tilde{\chi}_n^i}{r_n} \right)_{t \in [0,T]} \right) \right] \right|
$$
$$
\leq ||H||_\infty \left(\lceil nTq_n \rceil \mathbb{P}\left(A_2^{(0)}(n, T)^c \right) + \mathbf{P}\left(\mathcal{D}(n, T)^c \right) \right)
$$

which converges to 0 as $n \to \infty$ for $\overline{C}$ large by the same argument as (4.3) and that $\mathbf{P}(\mathcal{D}(n, T)^c) \to 0$. □

Using Corollary 5.5 and Lemma 5.6, in order to show the convergence of $\Delta_{\lfloor nt \rfloor}/r_n$, it suffices to show the convergence of the scaled sum of independent random variables $\tilde{\chi}_{t,n}/r_n$.

6 Excursion times in dense branches

In this section we only consider IVFE. The main tool will be Theorem 6, which is Theorem 10.2 in [4], and is itself a consequence of Theorem IV.6 in [23].

Theorem 6 *Let $n(t) : [0, \infty) \to \mathbb{N}$ and for each t let $\{R_k(t)\}_{k=1}^{n(t)}$ be a sequence of i.i.d. random variables. Assume that for every $\epsilon > 0$ it is true that*

$$\lim_{t \to \infty} \mathbb{P}(R_1(t) > \epsilon) = 0.$$

Now let $\mathcal{L}(x) : \mathbb{R}\backslash\{0\} \to \mathbb{R}$ be a real, non-decreasing function satisfying $\lim_{x \to \infty} \mathcal{L}(x) = 0$ and $\int_0^a x^2 d\mathcal{L}(x) < \infty$ for all $a > 0$. Suppose $d \in \mathbb{R}$ and $\sigma \geq 0$, then the following statements are equivalent:

1. *As $t \to \infty$*

$$\sum_{k=1}^{n(t)} R_k(t) \overset{d}{\to} R_{d,\sigma,\mathcal{L}}$$

   where $R_{d,\sigma,\mathcal{L}}$ has the law $\mathcal{I}(d, \sigma, \mathcal{L})$, that is,

$$\mathbb{E}\left[e^{it R_{d,\sigma,\mathcal{L}}}\right] := \exp\left(idt + \int_0^\infty \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) d\mathcal{L}(x)\right). \quad (6.1)$$

2. *For $\tau > 0$ let $\overline{R}_\tau(t) := R_1(t)\mathbf{1}_{\{|R_1(t)| \leq \tau\}}$ then for every continuity point x of $\mathcal{L}$*

$$d = \lim_{t \to \infty} n(t)\mathbb{E}[\overline{R}_\tau(t)] + \int_{|x| > \tau} \frac{x}{1+x^2} d\mathcal{L}(x) - \int_{0 < |x| \leq \tau} \frac{x^3}{1+x^2} d\mathcal{L}(x),$$

$$\sigma^2 = \lim_{\tau \to 0} \limsup_{t \to \infty} n(t) Var(\overline{R}_\tau(t)),$$

$$\mathcal{L}(x) = \begin{cases} \lim_{t \to \infty} n(t)\mathbb{P}(R_1(t) \leq x) & x < 0 \\ -\lim_{t \to \infty} n(t)\mathbb{P}(R_1(t) > x) & x > 0 \end{cases}$$

In our case, $n(t)$ will be the number of large branches up to level $\lfloor nt \rfloor$ and $\{R_k\}_{k=1}^{n(t)}$ independent copies of the time spent in a large branch.

Since we are now working with i.i.d. random variables we will simplify notation by considering the dummy branch $\mathcal{T}^{*-}$ defined in Definition 4 which has root ρ and first generation vertices $\rho_1, \ldots, \rho_{\xi^*-1}$ which are roots of f-GW-trees $(\mathcal{T}_j^\circ)_{j=1}^{\xi^*-1}$ (Fig. 3). We then let $(W^j)_{j=1}^{\xi^*-1}$ have the multinomial distribution determined in Lemma 5.1; that is, W^j represents the number of excursions into the jth trap of $\mathcal{T}^{*-}$. For the biased random walk X_n on $\mathcal{T}^{*-}$ started from ρ, let $T^{j,k}$ denote the duration of the kth excursion in the jth trap where we recall that in IVFE the excursion starts and ends at the root ρ. We then have that
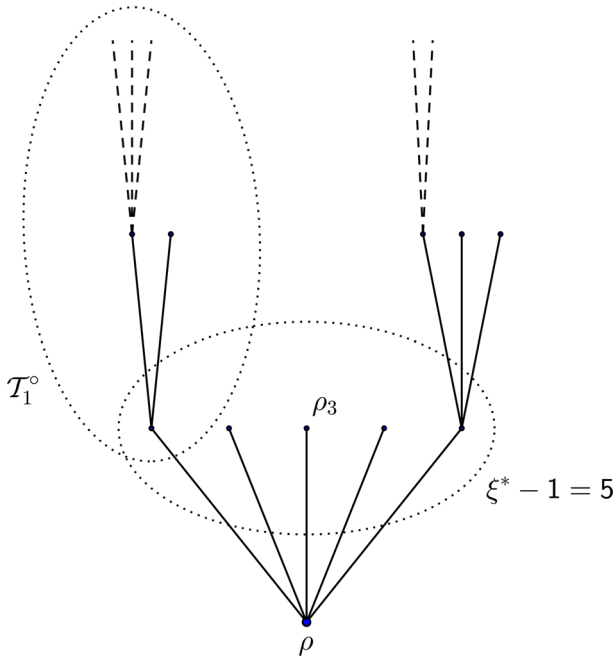
Fig. 3 A dummy tree $\mathcal{T}^{*-}$ with five buds, each of which is the root of an independent, unconditioned subcritical GW-tree

$$\tilde{\chi}_n := \sum_{j=1}^{\xi^*-1} \sum_{k=1}^{W^j} T^{j,k}. \tag{6.2}$$

is equal in distribution under $\mathbb{P}(\cdot|\xi^* > l_{n,\varepsilon})$ to $\tilde{\chi}_n^i$ under $\mathbb{P}$ for any i.

For $K \geq l_{n,\varepsilon} - l_{n,0}$ write $\overline{L}_K := l_{n,0} + K$ then denote $\mathbb{P}^K(\cdot) := \mathbb{P}\left(\cdot|\xi^* - 1 = \overline{L}_K\right)$ and $\mathbf{P}^K(\cdot) := \mathbf{P}\left(\cdot|\xi^* - 1 = \overline{L}_K\right)$. We now proceed to show that under $\mathbb{P}^K$

$$\zeta^{(n)} := \frac{1}{\xi^* - 1} \sum_{j=1}^{\xi^*-1} \sum_{k=1}^{W^j} T^{j,k} \tag{6.3}$$

converges in distribution to some random variable Z_∞ whose distribution doesn't depend on K.

We start by showing that the excursion times $T^{j,k}$ don't differ greatly from $E^{\mathcal{T}^{*-}}[T^{j,k}]$. In order to do this we require moment bounds on $T^{j,k}$ however since $\mathbf{E}[\xi^2] = \infty$ we don't have finite variance of the excursion times and thus we require a more subtle treatment. Recall that for a tree T we denote Z_n^T to be the size of the nth generation. Excursion times are first return times τ_ρ^+ conditioned on the first step therefore pruning buds and using (5.1) we have that the expected excursion time in a trap $\mathcal{T}_j^\circ$ is

$$E_\rho^{T^{*-}}\left[T^{j,k}\right] = E_\rho^{T^{*-}}\left[\tau_\rho^+ | X_1 = \rho_j\right] = \sum_{n=0}^\infty Z_n^{T_j^\circ} \beta^n \leq \mathcal{H}\left(T_j^\circ\right) \sup_n Z_n^{T_j^\circ} \beta^n. \quad (6.4)$$

Using that $\mathbf{P}(Z_n^{T^\circ} > 0) \sim c_\mu \mu^n$ (from (3.1)) we see that for n large there are no traps of height greater than $C \log(n)$ for some constant C thus for our purposes it will suffice to study $\sup_n Z_n^{T^\circ} \beta^n$.

Lemma 6.1 *Let Z_n be a subcritical Galton–Watson process with mean μ and offspring ξ satisfying $\mathbf{E}[\xi^{1+\tilde\varepsilon}] < \infty$ for some $\tilde\varepsilon > 0$. Suppose $1 < \beta < \mu^{-1}$, then there exists $\kappa > 0$ such that for all $\epsilon \in (0, \kappa)$ we have that $(Z_n \beta^n)^{1+\epsilon}$ is a supermartingale.*

Proof Let $\mathcal{F}_n := \sigma(Z_k; \ k \leq n)$ denote the natural filtration of Z_n and $(\xi_k)_{k\geq1}$ be independent copies of ξ.

$$\mathbf{E}\left[(Z_n\beta^n)^{1+\epsilon} | \mathcal{F}_{n-1}\right] = \left(Z_{n-1}\beta^{n-1}\right)^{1+\epsilon} \beta^{1+\epsilon} \mathbf{E}\left[\left(\sum_{k=1}^{Z_{n-1}} \frac{\xi_k}{Z_{n-1}}\right)^{1+\epsilon} \Big| Z_{n-1}\right]$$

$$\leq \left(Z_{n-1}\beta^{n-1}\right)^{1+\epsilon} \beta^{1+\epsilon} \mathbf{E}\left[\sum_{k=1}^{Z_{n-1}} \frac{\xi_k^{1+\epsilon}}{Z_{n-1}} \Big| Z_{n-1}\right]$$

$$= \left(Z_{n-1}\beta^{n-1}\right)^{1+\epsilon} \beta^{1+\epsilon} \mathbf{E}\left[\xi^{1+\epsilon}\right]$$

where the inequality follows by convexity of $f(x) = x^{1+\epsilon}$. From this it follows that for $\epsilon \in (0, \alpha - 1)$

$$\mathbf{E}\left[(Z_n\beta)^{1+\epsilon}\right] \leq \mathbf{E}\left[(Z_{n-1}\beta)^{1+\epsilon}\right]\mathbf{E}\left[(\xi\beta)^{1+\epsilon}\right] \leq \mathbf{E}\left[(\xi\beta)^{1+\epsilon}\right]^n < \infty.$$

Fix $\lambda = (\mu/\beta)^{1/2}$ then $\mu < \lambda$ and for $\epsilon > 0$ sufficiently small $\lambda\beta^{1+\epsilon} < 1$. By dominated convergence $\mathbf{E}[\xi^{1+\epsilon}] < \lambda$ for all ϵ small. In particular, $\beta^{1+\epsilon}\mathbf{E}[\xi^{1+\epsilon}] < 1$ for ϵ suitably small and therefore $(Z_n\beta^n)^{1+\epsilon}$ is a supermartingale. □

Lemma 6.2 *In IVFE, we can choose $\varepsilon > 0$ such that for any $t > 0$ there exists a constant C_t such that*

$$\sup_{K \geq -(a_n - l_{n,\varepsilon})} \mathbb{P}^K\left(\left|\frac{1}{\overline{L}_K} \sum_{j=1}^{\overline{L}_K} \sum_{k=1}^{W^j} (T^{j,k} - E^{T^{*-}}\left[T^{j,1}\right])\right| > t\right) \leq C_t n^{-2\varepsilon}.$$

Proof Write $E_m := \bigcap_{j=1}^m \left\{\mathcal{H}(T_j^\circ) \leq C \log(m)\right\}$ to be the event that none of the first m trees have height greater than $C \log(m)$. From (3.1) $\mathbf{P}(\mathcal{H}(T_j^\circ) \geq m) \sim c_\mu \mu^m$ hence we can choose $\tilde{C} > c_\mu$ such that

$$\mathbf{P}(E_m^c) \leq m\mathbf{P}\left(\mathcal{H}(T_j^\circ) > C \log(m)\right) \leq \tilde{C}m\mu^{C \log(m)}.$$

Thus choosing $C > 1/\log(\mu^{-1})$ and $c = C\log(\mu^{-1}) - 1 > 0$ we have that $\mathbf{P}(E_m^c) \leq \tilde{C}m^{-c}$ for m sufficiently large. By Lemma 6.1 we have that $(Z_k\beta^k)^{1+\epsilon}$ is a supermartingale for $\epsilon > 0$ sufficiently small where Z_n is the process associated to $\mathcal{T}^\circ$ therefore by Doob's supermartingale inequality

$$\mathbf{P}\left(\sup_{k\leq m} Z_k\beta^k \geq x\right) = \mathbf{P}\left(\sup_{k\leq m}(Z_k\beta^k)^{1+\epsilon} \geq x^{1+\epsilon}\right) \leq \mathbf{E}\left[Z_0^{1+\epsilon}\right]x^{-(1+\epsilon)}.$$

Using the expression (6.4) for the expected excursion time it follows that

$$\mathbf{P}\left(E^{\mathcal{T}_j^\circ}\left[T^{j,1}\right] > x \middle| \mathcal{H}\left(\mathcal{T}_j^\circ\right) \leq C\log(m)\right) \leq C\log(m)^{1+\epsilon}x^{-(1+\epsilon)}. \tag{6.5}$$

In particular, for some slowly varying function $\overline{L}$

$$\mathbf{E}\left[E^{\mathcal{T}_j^\circ}\left[T^{j,1}\right]^2 \mathbf{1}_{\left\{E^{\mathcal{T}_j^\circ}[T^{j,1}]\leq m\right\}} \middle| \mathcal{H}\left(\mathcal{T}_j^\circ\right) \leq C\log(m)\right] \leq C\overline{L}(m)m^{1-\epsilon}. \tag{6.6}$$

Let $\kappa = \epsilon/(2(1+\epsilon))$ then write $\overline{E}_m := E_m \cap \bigcap_{j=1}^m\{E^{\mathcal{T}_j^\circ}[T^{j,1}] \leq m^{1-\kappa}\}$ to be the event that no trap is of height greater than $C\log(m)$ and the expected excursion time in any trap is at most $m^{1-\kappa}$. For m sufficiently large by (6.5) we have that

$$\mathbf{P}\left(\overline{E}_m^c\right) \leq \mathbf{P}\left(\bigcup_{j=1}^m\left\{E^{\mathcal{T}_j^\circ}\left[T^{j,1}\right] > m^{1-\kappa}\right\}\middle|\mathcal{H}(\mathcal{T}_j^\circ) \leq C\log(m)\ \forall j \leq m\right) + \mathbf{P}\left(E_m^c\right)$$

$$\leq mC\log(m)^{1+\epsilon}m^{-(1-\kappa)(1+\epsilon)} + \tilde{C}m^{-c}.$$

Write $\overline{\overline{E}}_m := \overline{E}_m \cap \bigcap_{j=1}^m\{W^j \leq C'\log(m)\}$ for $C' > (2\beta-1)/(\beta-1)$ to be the event that no trap is of height greater than $C\log(m)$, entered more than $C'\log(m)$ times or has expected excursion time greater than $m^{1-\kappa}$. Then, by a union bound and the geometric distribution of W^j from Lemma 5.1

$$\mathbb{P}\left(\overline{\overline{E}}_m^c\right) \leq \mathbf{P}(\overline{E}_m^c) + m\mathbb{P}(W^1 > C'\log(m))$$

$$\leq \tilde{C}\left(\log(m)^{1+\epsilon}m^{1-(1-\kappa)(1+\epsilon)} + m^{-c} + m^{1-C'\frac{\beta-1}{2\beta-1}}\right) \tag{6.7}$$

for m large. Since $(1 - \kappa)(1 + \epsilon) > 1$ we can choose $\varepsilon < \frac{1}{2} \min \{(1 - \kappa)(1 + \epsilon) - 1,$ $c, C' \frac{\beta - 1}{2\beta - 1} - 1\}$ then we have that $\mathbb{P}\left(\overline{\overline{E}}_m^c\right) \leq \tilde{C} m^{-2\varepsilon}$ and

$$
\mathbb{P}\left(\left|\frac{1}{m} \sum_{j=1}^{m} \sum_{k=1}^{W^j} (T^{j,k} - E^{T_j^\circ}[T_{j,k}])\right| > t\right)
$$

$$
\leq \mathbb{E}\left[\frac{\sum_{j=1}^{m} C \log(m) Var_{P^{T_j^\circ}}\left(\left(T^{j,1} - E^{T_j^\circ}[T^{j,1}]\right) \mathbb{1}_{\overline{\overline{E}}_m}\right)}{(mt)^2}\right] + \mathbb{P}\left(\overline{\overline{E}}_m^c\right)
$$

$$
\leq \frac{C \log(m)}{mt^2} m^{(1-\epsilon)} \overline{L}(m) + \tilde{C} m^{-2\varepsilon}
$$

for some slowly varying function $\overline{L}$. Here the first inequality comes from Chebyshev and the second holds due to (6.6). Since $\epsilon > 0$ we can choose $\varepsilon \in (0, \epsilon/2)$ then

$$
\mathbb{P}\left(\left|\frac{1}{m} \sum_{j=1}^{m} \sum_{k=1}^{W^j} \left(T_{j,k} - E^{T_j^\circ}[T_{j,k}]\right)\right| > t\right) \leq C_t m^{-2\varepsilon}.
$$

In particular, this holds for $m = \overline{L}_K \geq a_{n^{1-\varepsilon}}$ thus since $\alpha < 2$

$$
\sup_{K \geq -(a_n - l_{n,\varepsilon})} \mathbb{P}^K\left(\left|\frac{1}{\overline{L}_K} \sum_{j=1}^{\overline{L}_K} \sum_{k=1}^{W^j} (T^{j,k} - E^{T_j^\circ}[T^{j,1}])\right| > t\right) \leq C_t \sup_{K \geq -(a_n - l_{n,\varepsilon})} \overline{L}_K^{-2\varepsilon}
$$

$$
\leq C_t a_{n^{1-\varepsilon}}^{-2\varepsilon}
$$

$$
= C_t n^{-2\varepsilon} \left(n^{\frac{2-\alpha}{\alpha-1} - \varepsilon} \tilde{L}(n^{1-\varepsilon})\right)^{-2\varepsilon}
$$

which is bounded above by $C_t n^{-2\varepsilon}$ for n large whenever $\varepsilon < (2 - \alpha)/(\alpha - 1)$. □

Using this we can now show that the average time spent in a trap indeed converges to its expectation.

Lemma 6.3 *In IVFE, we can find $\varepsilon > 0$ such that for sufficiently large n we have that*

$$
\sup_{K \geq -(a_n - l_{n,\varepsilon})} \mathbb{P}^K\left(\left|\frac{1}{\overline{L}_K} \sum_{j=1}^{\overline{L}_K} W^j \left(E^{T_j^\circ}[T^{j,1}] - \mathbb{E}[T^{1,1}]\right)\right| > t\right) \leq r(n)\left(n^{-\varepsilon} + \frac{C}{t}\right)
$$

uniformly over $t \geq 0$ where $r(n) = o(1)$.

Proof We continue using the notation defined in Lemma 6.2 and also define the event

$$
E_m^j := \left\{\mathcal{H}\left(T_j^\circ\right) \leq \tilde{C} \log(m)\right\} \cap \left\{W^j \leq C \log(m)\right\} \cap \left\{E^{T_j^\circ}[T^{j,1}] \leq m^{1-\kappa}\right\}
$$

that the jth trap isn't tall, entered many times and that the expected excursion time in it isn't large.

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{j=1}^{m}W^j\left(E^{\mathcal{T}_j^\circ}\left[T^{j,1}\right]-\mathbb{E}\left[T^{1,1}\right]\right)\right|>t\right)$$

$$\leq\mathbb{E}\left[\mathbb{P}\left(\left|\frac{1}{m}\sum_{j=1}^{m}W^j\left(E^{\mathcal{T}_j^\circ}\left[T^{j,1}\right]\mathbf{1}_{E_m^j}-\mathbb{E}\left[T^{1,1}\right]\mathbf{1}_{E_m^j}\right)\right|>t\left|(W^j)_{j=1}^m\right.\right)\right]+o(m^{-\varepsilon}).$$

Since $\mathbb{E}[E^{\mathcal{T}_j^\circ}[T^{j,1}\mathbf{1}_{E_m^j}]]=\mathbb{E}[T^{j,1}\mathbf{1}_{E_m^j}]\neq\mathbb{E}[\mathbb{E}[T^{1,1}]\mathbf{1}_{E_m^j}]$ we have that the summand in the right hand side doesn't have zero mean thus we perform the splitting:

$$\mathbb{E}\left[\mathbb{P}\left(\left|\frac{1}{m}\sum_{j=1}^{m}W^j(E^{\mathcal{T}_j^\circ}[T^{j,1}]\mathbf{1}_{E_m^j}-\mathbb{E}[T^{j,1}]\mathbf{1}_{E_m^j})\right|>t\left|(W^j)_{j=1}^m\right.\right)\right]$$

$$\leq\mathbb{E}\left[\mathbb{P}\left(\left|\frac{1}{m}\sum_{j=1}^{m}W^j(E^{\mathcal{T}_j^\circ}[T^{j,1}]\mathbf{1}_{E_m^j}-\mathbb{E}[T^{j,1}\mathbf{1}_{E_m^j}])\right|>t/3\left|(W^j)_{j=1}^m\right.\right)\right]$$

$$+\mathbb{E}\left[\mathbb{P}\left(\left|\frac{1}{m}\sum_{j=1}^{m}W^j(\mathbb{E}[T^{j,1}\mathbf{1}_{E_m^j}]-\mathbb{E}[T^{j,1}\mathbf{1}_{E_m^j}]\mathbf{1}_{E_m^j})\right|>t/3\left|(W^j)_{j=1}^m\right.\right)\right]$$

$$+\mathbb{E}\left[\mathbb{P}\left(\left|\frac{1}{m}\sum_{j=1}^{m}W^j(\mathbb{E}[T^{j,1}]\mathbf{1}_{E_m^j}-\mathbb{E}[T^{j,1}\mathbf{1}_{E_m^j}]\mathbf{1}_{E_m^j})\right|>t/3\left|(W^j)_{j=1}^m\right.\right)\right].$$

By Chebyshev's inequality and the tail bound $\mathbb{E}[E^{\mathcal{T}_j^\circ}[T^{j,1}]^2\mathbf{1}_{\{E_m^j\}}]\leq Cm^{1-\epsilon}L(m)$ from (6.6) we have that the first term is bounded above by

$$\mathbb{E}\left[\frac{C\log(m)^2}{(mt/3)^2}\sum_{j=1}^{m}Var\left(E^{\mathcal{T}_j^\circ}\left[T^{j,1}\right]\mathbf{1}_{E_m^j}\right)\right]\leq C_t m^{-\epsilon}\overline{L}(m)$$

for some slowly varying function $\overline{L}$. The second term is equal to

$$\mathbb{E}\left[\mathbb{P}\left(\left|\frac{1}{m}\sum_{j=1}^{m}W^j\mathbb{E}[T^{1,1}\mathbf{1}_{E_m^j}]\mathbf{1}_{(E_m^j)^c}\right|>t/3\left|(W^j)_{j=1}^m\right.\right)\right]$$

$$\leq\mathbf{P}\left(\bigcup_{j=1}^{m}(E_m^j)^c\right)=o(m^{-\varepsilon})$$

by (6.7). The final term can be written as

$$\mathbb{P}\left(\frac{1}{m}\sum_{j=1}^{m} W^j \mathbb{E}[T^{j,1}\mathbf{1}_{(E_m^j)^c}]\mathbf{1}_{E_m^j} > t/3\right) \leq \frac{3}{mt}\sum_{j=1}^{m} \mathbb{E}[W^j]\mathbb{E}[T^{j,1}\mathbf{1}_{(E_m^j)^c}]$$

$$= \frac{C}{t}\mathbb{E}[T^{1,1}\mathbf{1}_{(E_m^1)^c}]$$

which converges to 0 as $m \to \infty$ by dominated convergence since, by (5.2), $\mathbb{E}[T^{1,1}] < \infty$. We therefore have that the statement holds by setting $m = \overline{L}_K$. □

Recall from (6.3) that, under $\mathbb{P}^K$, $\zeta^{(n)}$ is the average time spent in a trap of a branch with $\xi^* - 1 = \overline{L}_K$ buds. From Lemmas 6.2 and 6.3 we have that as $n \to \infty$

$$\sup_{K \geq -(a_n - l_{n,\varepsilon})} \mathbb{P}^K\left(\left|\zeta^{(n)} - \mathbb{E}[T^{1,1}]\sum_{j=1}^{\overline{L}_K} \frac{W^j}{\overline{L}_K}\right| > t\right) \to 0.$$

Using (5.1) we have that $\mathbb{E}[T^{1,1}] = 2/(1 - \beta\mu)$. Write $\theta = (\beta - 1)(1 - \beta\mu)/(2\beta)$ and let $Z^\infty \sim \exp(\theta)$.

Corollary 6.4 *In IVFE, we can find $\varepsilon > 0$ such that for sufficiently large n we have that*

$$\sup_{K \geq -(a_n - l_{n,\varepsilon})} \left|\mathbb{P}^K\left(\zeta^{(n)} > t\right) - \mathbb{P}\left(Z^\infty > t\right)\right| \leq \tilde{r}(n)\left(n^{-\varepsilon} + \frac{C}{t}\right)$$

uniformly over $t \geq 0$ where $\tilde{r}(n) = o(1)$.

Proof By Lemma 5.1 the sum of W^j have a geometric law. In particular,

$$\left|\mathbb{P}(Z^\infty > t) - \mathbb{P}^K\left(\mathbb{E}[T^{1,1}]\sum_{j=1}^{\overline{L}_K} \frac{W^j}{\overline{L}_K} > t\right)\right|$$

$$= \left|e^{-\theta t} - \mathbb{P}^K\left(Geo\left(\frac{\beta - 1}{(\overline{L}_K + 1)\beta - 1}\right) > \frac{\overline{L}_K t}{\mathbb{E}[T^{1,1}]}\right)\right|$$

$$= \left|e^{-\theta t} - \left(1 - \frac{\beta - 1}{(\overline{L}_K + 1)\beta - 1}\right)^{\left\lceil\frac{\overline{L}_K t}{\mathbb{E}[T^{1,1}]}\right\rceil}\right|$$

$$= \left|e^{-\theta t} - e^{-\theta t \frac{\overline{L}_K \beta}{\overline{L}_K \beta + \beta - 1}}\right| + o\left(\overline{L}_K^{-1}\right)$$

$$\leq Ce^{-\theta t}\overline{L}_K^{-1} + o\left(\overline{L}_K^{-1}\right)$$

for some constant C independent of K. It therefore follows that the laws of $\zeta^{(n)}$ converge under $\mathbb{P}^K$ to an exponential law. In particular, using Lemmas 6.2 and 6.3 with the bound

$$\left| \mathbb{P}^K \left(\zeta^{(n)} > t \right) - \mathbb{P} \left(\mathbb{E}[T^{1,1}] \sum_{j=1}^{\overline{L}_K} \frac{W^j}{\overline{L}_K} > t \right) \right|$$

$$\leq \mathbb{P}^K \left(\left| \frac{1}{\overline{L}_K} \sum_{j=1}^{\overline{L}_K} W^j \left(E^{T_j^\circ} \left[T^{j,1} \right] - \mathbb{E} \left[T^{1,1} \right] \right) \right| > \epsilon \right)$$

$$+ \mathbb{P} \left(Z^\infty \in [t - \epsilon, t + \epsilon] \right) + O \left(\overline{L}_K^{-1} \right)$$

with $\epsilon = r(n)^{1/2} t$, we have the result since $\overline{L}_K \geq l_{n,\varepsilon} \gg n^\varepsilon$. $\qquad\square$

Corollary 6.5 *In IVFE, for any $\tau > 0$ fixed*

$$\lim_{n \to \infty} \sup_{C \geq 0} \sup_{K \geq -(a_n - l_{n,\varepsilon})} (C \vee 1) \left| \mathbb{E} \left[Z^\infty \mathbf{1}_{\{C Z^\infty \leq \tau\}} \right] - \mathbb{E}^K \left[\zeta^{(n)} \mathbf{1}_{\{C \zeta^{(n)} \leq \tau\}} \right] \right| = 0.$$

Lemma 6.6 shows that the product of an exponential random variable with a heavy tailed random variable has a similar tail to the heavy tailed variable.

Lemma 6.6 *Let $X \sim exp(\theta)$ and ξ be an independent variable which belongs to the domain of attraction of a stable law of index $\alpha \in (0, 2)$. Then $\mathbf{P}(X\xi > x) \sim \theta^{-\alpha} \Gamma(\alpha + 1) \mathbf{P}(\xi > x)$ as $x \to \infty$.*

Proof For some slowly varying function L we have that $\mathbf{P}(\xi \geq t) \sim x^{-\alpha} L(x)$ as $x \to \infty$.

Fix $0 < u < 1 < v < \infty$ then $\forall y \leq u$ we have that $x/y > x$ thus $\mathbf{P}(\xi \geq x/y) \leq \mathbf{P}(\xi \geq x)$ it therefore follows that

$$0 \leq \int_0^u \theta e^{-\theta y} \frac{\mathbf{P}(\xi \geq x/y)}{\mathbf{P}(\xi \geq x)} dy \leq \int_0^u \theta e^{-\theta y} dy = 1 - e^{\theta u}.$$

For $y \in [u, v]$ we have that $\mathbf{P}(\xi \geq x/y)/\mathbf{P}(\xi \geq x) \to y^\alpha$ uniformly over y therefore

$$\lim_{x \to \infty} \int_u^v \theta e^{-\theta y} \frac{\mathbf{P}(\xi \geq x/y)}{\mathbf{P}(\xi \geq x)} dy = \int_u^v \theta e^{-\theta y} y^\alpha dy.$$

Moreover, since this holds for all $u \geq 0$ and $1 - e^{\theta u} \to 0$ as $u \to 0$ we have that

$$\lim_{x \to \infty} \int_0^v \theta e^{-\theta y} \frac{\mathbf{P}(\xi \geq x/y)}{\mathbf{P}(\xi \geq x)} dy = \int_0^v \theta e^{-\theta y} y^\alpha dy. \tag{6.8}$$

Since $0 < \mathbf{P}(\xi \geq x) \leq 1$ for all $x < \infty$ we have that L is bounded away from $0, \infty$ on any compact interval thus satisfies the requirements of Potter's theorem (see

for example [6], 1.5.4) that if L is slowly varying and bounded away from $0, \infty$ on any compact subset of $[0, \infty)$ then for any $\epsilon > 0$ there exists $A_\epsilon > 1$ such that for $x, y > 0$

$$\frac{L(z)}{L(x)} \leq A_\epsilon \max \left\{ \left(\frac{z}{x}\right)^\epsilon, \left(\frac{x}{z}\right)^\epsilon \right\}.$$

Moreover, $\exists c_1, c_2 > 0$ such that $c_1 t^{-\alpha} L(t) \leq \mathbf{P}(\xi \geq t) \leq c_2 t^{-\alpha} L(t)$ hence we have that for all $y > v$ $\mathbf{P}(\xi \geq x/y)/\mathbf{P}(\xi \geq x) \leq C y^{\alpha + \epsilon}$. By dominated convergence we therefore have that

$$\lim_{x \to \infty} \int_v^\infty \theta e^{-\theta y} \frac{\mathbf{P}(\xi \geq x/y)}{\mathbf{P}(\xi \geq x)} dy = \int_v^\infty \theta e^{-\theta y} y^\alpha dy.$$

Combining this with (6.8) we have that

$$\lim_{x \to \infty} \frac{\mathbf{P}(X\xi \geq x)}{\mathbf{P}(\xi \geq x)} = \lim_{x \to \infty} \int_0^\infty \theta e^{-\theta y} \frac{\mathbf{P}(\xi \geq x/y)}{\mathbf{P}(\xi \geq x)} dy$$
$$= \int_0^\infty \theta e^{-\theta y} y^\alpha dy = \theta^{-\alpha} \Gamma(\alpha + 1).$$

$\square$

We write $\mathbb{P}^>(\cdot) := \mathbb{P}(\cdot | \xi^* > l_{n,\varepsilon})$ and $\mathbf{P}^>(\cdot) := \mathbf{P}(\cdot | \xi^* > l_{n,\varepsilon})$ to be the laws conditioned on the branch $\mathcal{T}^{*-}$ being large. From (6.2) we have that (under $\mathbb{P}$) $\tilde{\chi}_n^i$ are independent copies of the time spent in a branch $\tilde{\chi}_n$ with respect to $\mathbb{P}^>$. Define $\tilde{\chi}_n^\infty := (\xi^* - 1) Z^\infty$ where Z^∞ is the exponential random variable used in Corollaries 6.4 and 6.5. Recall that $R_{d,\sigma,\mathcal{L}}$ has the infinitely divisible given by (6.1). Fix the sequence $(\lambda_n)_{n \geq 1}$ converging to some $\lambda > 0$ and denote $M_n^\lambda := \lfloor \lambda_n n^\varepsilon \rfloor$.

Proposition 6.7 *In IVFE, for any $\lambda > 0$, as $n \to \infty$*

$$\sum_{i=1}^{M_n^\lambda} \frac{\tilde{\chi}_n^i}{a_n} \xrightarrow{d} R_{d_\lambda, 0, \mathcal{L}_\lambda}$$

where

$$d_\lambda = \int_0^\infty \frac{x}{1 + x^2} d\mathcal{L}_\lambda(x)$$

$$\mathcal{L}_\lambda(x) = \begin{cases} 0 & x < 0 \\ -\lambda x^{-(\alpha-1)} \theta^{-(\alpha-1)} \Gamma(\alpha) & x > 0. \end{cases}$$

Proof Let $\epsilon > 0$ then by Markov's inequality

$$\mathbb{P}^> \left(\frac{\tilde{\chi}_n}{a_n} > \epsilon \right) \leq \mathbb{P}^> (\xi^* - 1 \geq a_n^{1-\varepsilon/2}) + \mathbb{P} \left(\sum_{j=1}^{a_n^{1-\varepsilon/2}} \sum_{k=1}^{W^j} T^{j,k} \geq \epsilon a_n \right)$$

$$\leq \frac{\mathbb{P}(\xi^* - 1 \geq a_n^{1-\varepsilon/2})}{\mathbb{P}(\xi^* - 1 \geq a_n^{1-\varepsilon})} + \frac{a_n^{1-\varepsilon/2}}{\epsilon a_n} \mathbb{E}[W^1] \mathbb{E}[T^{1,1}],$$

which converges to 0 as $n \to \infty$. Thus, by Theorem 6, it suffices to show that

1.
$$\lim_{\tau \to 0^+} \limsup_{n \to \infty} M_n^\lambda Var_{\mathbb{P}^>} \left(\frac{\tilde{\chi}_n}{a_n} \mathbf{1}_{\{\frac{\tilde{\chi}_n}{a_n} \leq \tau\}} \right) = 0,$$

2.
$$\mathcal{L}_\lambda(x) = \begin{cases} \lim_{n \to \infty} M_n^\lambda \mathbb{P}^> \left(\frac{\tilde{\chi}_n}{a_n} \leq x \right) & x < 0, \\ -\lim_{n \to \infty} M_n^\lambda \mathbb{P}^> \left(\frac{\tilde{\chi}_n}{a_n} > x \right) & x > 0, \end{cases}$$

3.
$$d_\lambda = \lim_{n \to \infty} M_n^\lambda \mathbb{E}^> \left[\frac{\tilde{\chi}_n}{a_n} \mathbf{1}_{\{\frac{\tilde{\chi}_n}{a_n} \leq \tau\}} \right] + \int_{|x| > \tau} \frac{x}{1+x^2} d\mathcal{L}_\lambda(x)$$
$$- \int_{0 < |x| \leq \tau} \frac{x^3}{1+x^2} d\mathcal{L}_\lambda(x)$$

where d_λ and $\mathcal{L}_\lambda$ are as stated above.

We start with the first condition and since $\lambda_n \to \lambda$ there exists a constant C such that

$$M_n^\lambda Var_{\mathbb{P}^>} \left(\frac{\tilde{\chi}_n}{a_n} \mathbf{1}_{\{\frac{\tilde{\chi}_n}{a_n} \leq \tau\}} \right) \leq Cn^\varepsilon \mathbb{E}^> \left[\left(\frac{\tilde{\chi}_n}{a_n} \right)^2 \mathbf{1}_{\{\frac{\tilde{\chi}_n}{a_n} \leq \tau\}} \right]$$
$$\leq Cn^\varepsilon \left(\tau^2 \mathbf{P}^> (\xi^* - 1 \geq a_n) + \tau \mathbb{E}^> \left[\frac{\tilde{\chi}_n}{a_n} \mathbf{1}_{\{\xi^* - 1 < a_n\}} \right] \right). \tag{6.9}$$

By the definition of a_n we have that

$$\mathbf{P}^> (\xi^* - 1 \geq a_n) = \frac{\mathbf{P}(\xi^* \geq a_n)}{\mathbf{P}(\xi^* \geq a_n^{1-\varepsilon})} \sim n^{-\varepsilon}. \tag{6.10}$$

Conditional on the number of buds ξ^* we have that the number of excursions W^j into the jth trap are independent from the excursion times $T^{j,k}$ and both the number of excursions and the excursion times have finite mean hence

$$\mathbb{E}^> \left[\frac{\tilde{\chi}_n}{a_n} \mathbf{1}_{\{\xi^* - 1 < a_n\}} \right] = \sum_{r=a_{n^{1-\varepsilon}}}^{a_n - 1} \frac{\mathbf{P}(\xi^* - 1 = r)}{\mathbf{P}(\xi^* - 1 \geq a_{n^{1-\varepsilon}})} \mathbb{E} \left[\sum_{j=1}^{r} \sum_{k=1}^{W^j} \frac{T^{j,k}}{a_n} \middle| \xi^* - 1 = r \right]$$

$$\leq \frac{\mathbb{E}[W^1] \mathbb{E}[T^{1,1}]}{\mathbf{P}(\xi^* - 1 \geq a_{n^{1-\varepsilon}})} \mathbf{E} \left[\frac{\xi^* - 1}{a_n} \mathbf{1}_{\{\xi^* - 1 \leq a_n\}} \right]$$

$$\sim Cn^{-\varepsilon}$$

where the asymptotic holds as $n \to \infty$ by (5.4). In particular, by combining this with
(6.10) in (6.9) we have that $M_n^\lambda Var_{\mathbb{P}^>}(\frac{\tilde{\chi}_n}{a_n} \mathbf{1}_{\{\frac{\tilde{\chi}_n}{a_n} \leq \tau\}}) \leq C(\tau^2 + \tau)$ for some constant
C depending on λ hence, as $\tau \to 0^+$, we indeed have convergence to 0 and therefore
the first condition holds.

We now move on to the Lévy spectral function $\mathcal{L}_\lambda$. Clearly for $x < 0$ we have that
$\mathcal{L}_\lambda(x) = 0$ since $\tilde{\chi}_n$ is a positive random variable. It therefore suffices to consider
$x > 0$. By Corollary 6.4 we have that the scaled time spent in a large trap $\zeta^{(n)}$ [from
(6.3)] converges in distribution to an exponential random variable Z^∞ with parameter
θ (which is independent of K) therefore, since $M_n^\lambda \sim \lambda n^\varepsilon$ and $\tilde{\chi}_n^\infty = (\xi^* - 1)Z^\infty$ we
have that

$$M_n^\lambda \mathbb{P}^> \left(\frac{\tilde{\chi}_n^\infty}{a_n} > x \right) \sim \lambda n^\varepsilon \mathbb{P}^> \left((\xi^* - 1)Z^\infty > xa_n \right)$$

$$\sim \frac{\lambda}{\mathbf{P}(\xi^* - 1 \geq a_n)} \sum_{K \geq l_{n,\varepsilon} - l_{n,0}} \mathbf{P}\left(\xi^* - 1 = \overline{L}_K \right) \mathbb{P}\left(\overline{L}_K Z^\infty > xa_n \right)$$

$$= \lambda \frac{\mathbb{P}\left((\xi^* - 1)Z^\infty \geq xa_n \right)}{\mathbf{P}(\xi^* - 1 \geq a_n)} - \sum_{j=0}^{l_{n,\varepsilon} - 1} \frac{\lambda \mathbf{P}\left(\xi^* - 1 = j \right) \mathbb{P}\left(jZ^\infty > xa_n \right)}{\mathbf{P}(\xi^* - 1 \geq a_n)}$$

$$\sim \lambda \theta^{-(\alpha-1)} \Gamma(\alpha) x^{-(\alpha-1)}.$$

Where the final asymptotic holds by Lemma 6.6 and because

$$\sum_{j=0}^{l_{n,\varepsilon} - 1} \frac{\lambda \mathbf{P}\left(\xi^* - 1 = j \right) \mathbb{P}\left(jZ^\infty > xa_n \right)}{\mathbf{P}(\xi^* - 1 \geq a_n)} \leq \lambda \frac{\mathbb{P}\left(Z^\infty > xa_n / a_{n^{1-\varepsilon}} \right)}{\mathbf{P}(\xi^* - 1 \geq a_n)} = \lambda \frac{e^{-\theta x \frac{a_n}{l_{n,\varepsilon}}}}{\mathbf{P}(\xi^* - 1 \geq a_n)}$$

which converges to 0 as $n \to \infty$ since $l_{n,\varepsilon} = a_{\lfloor n^{1-\varepsilon} \rfloor}$ (and therefore $a_n / l_{n,\varepsilon} >> n^\varepsilon$).
It now suffices to show that $n^\varepsilon \left(\mathbb{P}^> \left(\frac{\tilde{\chi}_n^\infty}{a_n} > x \right) - \mathbb{P}^> \left(\frac{\tilde{\chi}_n}{a_n} > x \right) \right)$ converges to 0 as
$n \to \infty$. To do this we condition on the number of buds:

$$\mathbb{P}^> \left(\frac{\tilde{\chi}_n^\infty}{a_n} > x \right) - \mathbb{P}^> \left(\frac{\tilde{\chi}_n}{a_n} > x \right)$$

$$= \sum_{K \geq l_{n,\varepsilon} - l_{n,0}} \mathbf{P}^> \left(\xi^* - 1 = \overline{L}_K \right) \left(\mathbb{P} \left(\frac{\overline{L}_K Z^\infty}{a_n} > x \right) - \mathbb{P}^K \left(\frac{\overline{L}_K \zeta^{(n)}}{a_n} > x \right) \right).$$

We consider positive and negative K separately. For $K \geq 0$ we have that

$$\sum_{K=0}^{\infty} n^{\varepsilon} \mathbf{P}^{>} \left(\xi^* - 1 = \overline{L}_K \right) \left| \mathbb{P} \left(\frac{\overline{L}_K Z^{\infty}}{a_n} > x \right) - \mathbb{P}^K \left(\frac{\overline{L}_K \zeta^{(n)}}{a_n} > x \right) \right|$$

$$\leq n^{\varepsilon} \mathbf{P}^{>} \left(\xi^* - 1 \geq a_n \right) \sup_{\substack{c \leq 1 \\ K \geq 0}} \left| \mathbb{P}^K (Z^{\infty} > cx) - \mathbb{P}^K (\zeta^{(n)} > cx) \right|. \qquad (6.11)$$

By (6.10) $n^{\varepsilon} \mathbf{P}^{>}(\xi^* - 1 \geq a_n)$ converges as $n \to \infty$ hence, using Corollary 6.4, (6.11) converges to 0. For $K \leq 0$, by Corollary 6.4 we have that

$$\sum_{K=-\infty}^{0} \mathbf{1}_{\{K \geq l_{n,\varepsilon} - l_{n,0}\}} n^{\varepsilon} \mathbf{P}^{>} \left(\xi^* - 1 = \overline{L}_K \right) \left| \mathbb{P} \left(\frac{\overline{L}_K Z^{\infty}}{a_n} > x \right) - \mathbb{P}^K \left(\frac{\overline{L}_K \zeta^{(n)}}{a_n} > x \right) \right|$$

$$\leq n^{\varepsilon} \sum_{K=-\infty}^{0} \mathbf{1}_{\{K \geq l_{n,\varepsilon} - l_{n,0}\}} \mathbf{P}^{>} \left(\xi^* - 1 = \overline{L}_K \right) \tilde{r}(n) \left(n^{-\varepsilon} + \frac{C_x \overline{L}_K}{a_n} \right)$$

$$\leq o(1) + \frac{C_x \tilde{r}(n) n^{\varepsilon}}{a_n} \sum_{K=-\infty}^{0} \mathbf{1}_{\{K \geq l_{n,\varepsilon} - l_{n,0}\}} \frac{\mathbf{P} \left(\xi^* - 1 = \overline{L}_K \right)}{\mathbf{P} \left(\xi^* - 1 \geq l_{n,\varepsilon} \right)} \overline{L}_K.$$

For some constant C we have that $\mathbf{P}(\xi^* - 1 \geq l_{n,\varepsilon}) \sim C n^{-(1-\varepsilon)}$ thus by (5.4)

$$\frac{C_x \tilde{r}(n) n^{\varepsilon}}{a_n} \sum_{K=-\infty}^{0} \mathbf{1}_{\{K \geq l_{n,\varepsilon} - l_{n,0}\}} \frac{\mathbf{P} \left(\xi^* - 1 = \overline{L}_K \right)}{\mathbf{P} \left(\xi^* - 1 \geq l_{n,\varepsilon} \right)} \overline{L}_K$$

$$\leq C_x \tilde{r}(n) n \mathbb{E} \left[\frac{\xi^* - 1}{a_n} \mathbf{1}_{\{\xi^* - 1 \leq a_n\}} \right] \sim C_x \tilde{r}(n).$$

In particular, since $\tilde{r}(n) = o(1)$, we indeed have that this converges to zero and thus we have the required convergence for $\mathcal{L}_{\lambda}$.

Finally, we consider the drift term d_{λ}. Since $\int_{0 < x \leq \tau} x \, d\mathcal{L}_{\lambda}(x) < \infty$ we have that

$$d_{\lambda} = \lim_{n \to \infty} M_n^{\lambda} \mathbb{E}^{>} \left[\frac{\tilde{\chi}_n}{a_n} \mathbf{1}_{\left\{ \frac{\tilde{\chi}_n}{a_n} \leq \tau \right\}} \right] + \int_0^{\infty} \frac{x}{1 + x^2} \, d\mathcal{L}_{\lambda}(x) - \int_0^{\tau} x \, d\mathcal{L}_{\lambda}(x).$$

We want to show that $d_{\lambda} = \int_0^{\infty} \frac{x}{1+x^2} d\mathcal{L}_{\lambda}(x)$ thus we need to show that the other terms cancel. By definition of $\mathbb{P}^{>}$ we have that

$$\mathbb{E}^{>} \left[\frac{\tilde{\chi}_n^{\infty}}{a_n} \mathbf{1}_{\left\{ \frac{\tilde{\chi}_n^{\infty}}{a_n} \leq \tau \right\}} \right] = \frac{1}{\mathbf{P} \left(\xi^* - 1 \geq l_{n,\varepsilon} \right)} \mathbb{E} \left[\frac{(\xi^* - 1) Z^{\infty}}{a_n} \mathbf{1}_{\left\{ \frac{(\xi^* - 1) Z^{\infty}}{a_n} \leq \tau \right\} \cap \{\xi^* > l_{n,\varepsilon}\}} \right].$$

By Lemma 6.6, $(\xi^* - 1) Z^{\infty}$ belongs to the domain of attraction of a stable law of index $\alpha - 1$ and satisfies the scaling properties of ξ^* (up to a constant factor). Therefore, using that $a_n \gg l_{n,\varepsilon}$, we have that

$$M_n^\lambda \mathbb{E}^> \left[\frac{\tilde{\chi}_n^\infty}{a_n} \mathbf{1}_{\left\{ \frac{\tilde{\chi}_n^\infty}{a_n} \leq \tau \right\}} \right] \sim \frac{\lambda n^\varepsilon}{\mathbf{P} \left(\xi^* - 1 \geq l_{n,\varepsilon} \right)} \mathbb{E} \left[\frac{(\xi^* - 1) Z^\infty}{a_n} \mathbf{1}_{\left\{ \frac{(\xi^* - 1) Z^\infty}{a_n} \leq \tau \right\}} \right]$$

$$\sim \frac{\alpha - 1}{2 - \alpha} \tau^{2 - \alpha} \lambda \theta^{-(\alpha - 1)} \Gamma(\alpha).$$

Using the form of the Lévy spectral function we have that

$$\int_0^\tau x \, d\mathcal{L}_\lambda(x) = \lambda \theta^{-(\alpha-1)} \Gamma(\alpha) \int_{\tau^{-(\alpha-1)}}^\infty x^{-\frac{1}{\alpha-1}} dx = \frac{\alpha - 1}{2 - \alpha} \tau^{2-\alpha} \lambda \theta^{-(\alpha-1)} \Gamma(\alpha)$$

thus it remains to show that

$$n^\varepsilon \left(\mathbb{E}^> \left[\frac{\tilde{\chi}_n^\infty}{a_n} \mathbf{1}_{\left\{ \frac{\tilde{\chi}_n^\infty}{a_n} \leq \tau \right\}} \right] - \mathbb{E}^> \left[\frac{\tilde{\chi}_n}{a_n} \mathbf{1}_{\left\{ \frac{\tilde{\chi}_n}{a_n} \leq \tau \right\}} \right] \right) \to 0.$$

Similarly to the previous parts we condition on $\xi^* - 1 = \overline{L}_K$ and consider the sums over K positive and negative separately. For $K \leq 0$

$$n^\varepsilon \sum_{K \leq 0} \mathbf{P}^> (\xi^* - 1 = \overline{L}_K) \left| \mathbb{E} \left[\frac{\overline{L}_K Z^\infty}{a_n} \mathbf{1}_{\left\{ \frac{\overline{L}_K Z^\infty}{a_n} \leq \tau \right\}} \right] - \mathbb{E}^K \left[\frac{\overline{L}_K \zeta^{(n)}}{a_n} \mathbf{1}_{\left\{ \frac{\overline{L}_K \zeta^{(n)}}{a_n} \leq \tau \right\}} \right] \right|$$

$$\leq \frac{n^\varepsilon}{\mathbf{P}(\xi^* \geq l_{n,\varepsilon})} \mathbb{E} \left[\frac{\xi^* - 1}{a_n} \mathbf{1}_{\{\xi^* \leq a_n\}} \right] \sup_{K \leq 0} \left| \mathbb{E} \left[Z^\infty \mathbf{1}_{\left\{ Z^\infty \leq \tau \frac{a_n}{\overline{L}_K} \right\}} \right] - \mathbb{E}^K \left[\zeta^{(n)} \mathbf{1}_{\left\{ \zeta^{(n)} \leq \tau \frac{a_n}{\overline{L}_K} \right\}} \right] \right|.$$

By definition of $l_{n,\varepsilon}$ and properties of stable laws $n^\varepsilon \mathbb{E} \left[(\xi^* - 1)/a_n \mathbf{1}_{\{\xi^* \leq a_n\}} \right] / \mathbf{P}(\xi^* \geq l_{n,\varepsilon})$ converges to some constant as $n \to \infty$. By Corollary 6.5 we therefore have that this converges to 0. Similarly for $K \geq 0$ we have that

$$n^\varepsilon \sum_{K \geq 0} \mathbf{P}^> (\xi^* - 1 = \overline{L}_K) \left| \mathbb{E} \left[\frac{\overline{L}_K Z^\infty}{a_n} \mathbf{1}_{\left\{ \frac{\overline{L}_K Z^\infty}{a_n} \leq \tau \right\}} \right] - \mathbb{E}^K \left[\frac{\overline{L}_K \zeta^{(n)}}{a_n} \mathbf{1}_{\left\{ \frac{\overline{L}_K \zeta^{(n)}}{a_n} \leq \tau \right\}} \right] \right|$$

$$\leq \frac{n^\varepsilon \mathbf{P}(\xi^* \geq l_{n,0})}{\mathbf{P}(\xi^* \geq l_{n,\varepsilon})} \sup_{K \geq 0} \frac{\overline{L}_K}{a_n} \left| \mathbb{E} \left[Z^\infty \mathbf{1}_{\left\{ Z^\infty \leq \tau \frac{a_n}{\overline{L}_K} \right\}} \right] - \mathbb{E}^K \left[\zeta^{(n)} \mathbf{1}_{\left\{ \zeta^{(n)} \leq \tau \frac{a_n}{\overline{L}_K} \right\}} \right] \right|.$$

We have that $n^\varepsilon \mathbf{P}(\xi^* \geq l_{n,0})/\mathbf{P}(\xi^* \geq l_{n,\varepsilon})$ converges to some constant as $n \to \infty$. The result then follows by Corollary 6.5. □

This shows the convergence result of Theorem 1 in the sense of finite dimensional distributions. In Sect. 9 we prove a tightness result which concludes the proof.

7 Excursion times in deep branches

In this section we decompose the time spent in large branches. In FVIE this will be very similar to the decomposition used in [4] and we won't consider the argument in great detail. However, the decomposition required in IVIE requires greater delicacy.

In Lemmas 7.1, 7.2 and Proposition 7.3 we consider a construction of a GW-tree conditioned on its height from [14] to show that the time spent in deep traps essentially consists of some geometric number of excursions from the deepest point in the trap to itself. That is, as in [4], excursions which don't reach the deepest point are negligible as is the time taken for the walk to reach the deepest point from the root of the trap and the time taken to return to the root from the deepest point when this happens before returning to the deepest point.

In the remainder of the section we show that, conditional on the exact height of the branch $\overline{H}$, the time spent in the branch scaled by $\beta^{\overline{H}}$ converges in distribution along the given subsequences. In Lemma 7.5 we determine an important asymptotic relation for the distribution over the number of buds conditional on the height of the branch. In Lemmas 7.6–7.9 we provide various bounds which allow us, in Proposition 7.10, to show that the excursion time in a large branch is close to the random variable Z_∞^n (defined in (7.23)) which removes some of the dependency on n.

The main result of the section is Proposition 7.14 which shows that the scaled time spent in a large branch converges in distribution along the given subsequences. As a prelude to this we prove Lemmas 7.11–7.13 which show that we can reintroduce small traps into the branch and that the height of a trap is sufficiently close to a geometric random variable. We then conclude the section by showing that the scaled excursion times can be dominated by some random variable with a certain moment property which will be important in Sect. 8.

Recall that $\mathcal{T}^\circ$ is an f-GW-tree and $\mathcal{H}(\mathcal{T}^\circ)$ is its height then, following notation of [4], we denote $(\phi_{n+1}, \psi_{n+1})_{n \geq 0}$ to be a sequence of i.i.d. pairs with joint law

$$
\begin{aligned}
&\mathbf{P}\left(\phi_{n+1} = j, \psi_{n+1} = k\right) \\
&\quad := \frac{\mathbf{P}(\xi = k)\mathbf{P}\left(\mathcal{H}\left(\mathcal{T}^\circ\right) \leq n - 1\right)^{j-1} \mathbf{P}\left(\mathcal{H}\left(\mathcal{T}^\circ\right) = n\right) \mathbf{P}\left(\mathcal{H}\left(\mathcal{T}^\circ\right) \leq n\right)^{k-j}}{\mathbf{P}\left(\mathcal{H}(\mathcal{T}^\circ) = n + 1\right)} \quad (7.1)
\end{aligned}
$$

for $k = 1, 2, \ldots$ and $j = 1, \ldots, k$. Under this law ψ_{n+1} has the law of the degree of the root of a GW-tree conditioned to be of height $n + 1$ and ϕ_{n+1} has the law over the first bud to give rise onto a tree of height exactly n. We then construct a sequence of trees recursively as follows: Set $\mathcal{T}_0^\prec = \{\delta\}$ then

1. Let the first generation of $\mathcal{T}_{n+1}^\prec$ be of size ψ_{n+1}.
2. Attach $\mathcal{T}_n^\prec$ to the ϕ_{n+1}th first generation vertex of $\mathcal{T}_{n+1}^\prec$.
3. Attach f-GW-trees conditioned to have height at most $n - 1$ to the first $\phi_{n+1} - 1$ vertices of the first generation of $\mathcal{T}_{n+1}^\prec$.
4. Attach f-GW-trees conditioned to have height at most n to the remaining $\psi_{n+1} - \phi_{n+1}$ first generation vertices of $\mathcal{T}_{n+1}^\prec$.

Under this construction $\mathcal{T}_{n+1}^\prec$ has the distribution of an f-GW-tree conditioned to have height exactly $n + 1$. Write $\delta_0 = \delta$ to be the deepest point of the tree and for $n = 1, 2, \ldots$ write δ_n to be the ancestor of δ of distance n. The sequence $\delta_0, \delta_1, \ldots$ form a 'spine' from the deepest point to the root of the tree. We denote $\mathcal{T}^\prec$ to be the tree asymptotically attained. By a subtrap of $\mathcal{T}^\prec$ we mean some vertex x on the spine together with a descendant y off the spine and all of the descendants of y. This is itself a tree with root x and we write $\mathcal{S}_x$ to be the collection of subtraps rooted at x. Figure 4
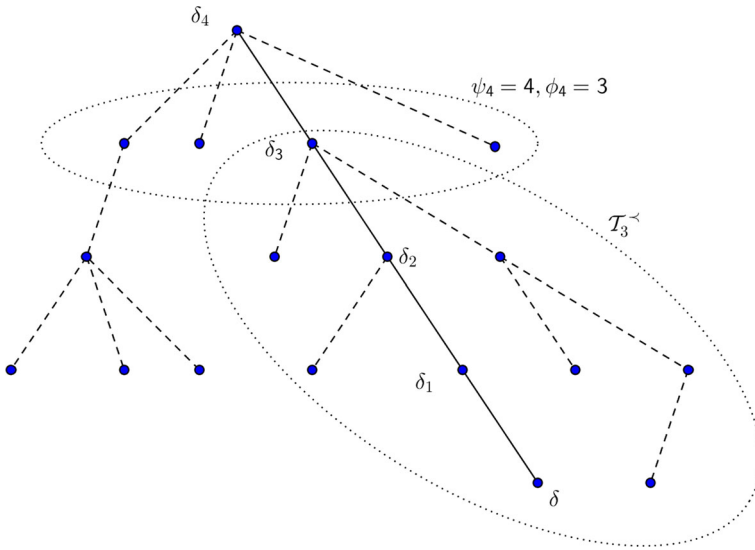
Fig. 4 A GW-tree conditioned to be of height 4 with the *solid line* representing the spine and *dashed lines* representing the subtraps which reach at most level 3 to the *left* of the spine and at most level 4 to the *right* of the spine

shows a construction of $\mathcal{T}_4^{\prec}$ where the solid line represents the spine and the dashed lines represent subtraps.

We denote $\mathcal{S}^{n,j,1}$ to be the jth subtrap conditioned to have height at most $n-1$ attached to δ_n and $\mathcal{S}^{n,j,2}$ to be the jth subtrap conditioned to have height at most n attached to δ_n. Recall that $d(x, y)$ denotes the graph distance between vertices x, y then for $k = 1, 2$ let

$$\Pi^{n,j,k} := 2 \sum_{x \in \mathcal{S}^{n,j,k} \setminus \{\delta_n\}} \beta^{d(x,\delta_n)}$$

denote the weight of $\mathcal{S}^{n,j,k}$ under the invariant measure associated to the conductance model with conductances β^{i+1} between levels $i, i+1$ and the roots of $\mathcal{S}^{n,j,k}$ (spinal vertices) denoting level 0. We then write

$$\Lambda_n := \sum_{j=1}^{\phi_n - 1} \Pi^{n,j,1} + \sum_{j=1}^{\psi_n - \phi_n} \Pi^{n,j,2}$$

to denote the total weight of the subtraps of δ_n then,

$$E^{\mathcal{T}^{\prec}}[\mathcal{R}_\infty] = 2 \sum_{n=0}^{\infty} \beta^{-n}(1 + \Lambda_n) \tag{7.2}$$

is the expected time $\mathcal{R}_\infty$ taken for a walk on $\mathcal{T}^{\prec}$ started from δ to return to δ.

Lemma 7.1 *Suppose that ξ belongs to the domain of attraction of a stable law of index $\alpha \in (1, 2]$ and $\beta\mu > 1$ then*

$$\mathbb{E}[\mathcal{R}_\infty] < \infty.$$

Proof Since $\beta > 1$ we have that $2\sum_{n=0}^{\infty} \beta^{-n} = 2/(1 - \beta^{-1})$ thus by (7.2) it suffices to find an appropriate bound on $\mathbf{E}[\Lambda_n]$.

$\mathbf{E}[\Pi^{n,j,1}] \leq \mathbf{E}[\Pi^{n,j,2}]$ since conditioning the height of the trap to be small only reduces the weight; therefore, by independence of ψ_n and $\Pi^{n,j,2}$

$$\mathbf{E}[\Lambda_n] = \mathbf{E}\left[\sum_{j=1}^{\phi_n-1} \Pi^{n,j,1} + \sum_{j=1}^{\psi_n-\phi_n} \Pi^{n,j,2}\right] \leq \mathbf{E}\left[\Pi^{n,1,2}\right]\mathbf{E}[\psi_n]. \qquad (7.3)$$

Using that conditioning the height of a GW-tree $\mathcal{T}^\circ$ to be small only decreases the expected generation sizes and that $\mu\beta > 1$, by (5.1)

$$\mathbf{E}\left[\Pi^{n,1,2}\right] = 2\sum_{k=1}^{n} \beta^k \mathbf{E}\left[Z_k \middle| \mathcal{H}(\mathcal{T}^\circ) \leq n\right] \leq c(\beta\mu)^n \qquad (7.4)$$

for some constant c where Z_k are the generation sizes of $\mathcal{T}^\circ$. Summing over j in (7.1) shows that $\mathbf{P}(\psi_{n+1} = k) = \mathbf{P}(Z_1 = k | \mathcal{H}(\mathcal{T}^\circ) = n + 1)$. Recalling that $s_n = \mathbf{P}(\mathcal{H}(\mathcal{T}^\circ) < n)$,

$$\mathbf{E}[\psi_{n+1}] = \mathbf{E}\left[Z_1 \middle| \mathcal{H}(\mathcal{T}^\circ) = n + 1\right] = \sum_{k=1}^{\infty} kp_k\left(\frac{s_{n+1}^k - s_n^k}{s_{n+2} - s_{n+1}}\right).$$

By (3.1) $1 - s_{n+1} \sim c\mu^n$ for some positive constant c. Let $\epsilon > 0$ be such that $1 - \epsilon - \mu(1+\epsilon) > 0$, then for n large we have that $(1-\epsilon)c\mu^n \leq 1 - s_{n+1} \leq (1+\epsilon)c\mu^n$. Therefore,

$$s_{n+2} - s_{n+1} = (1 - s_{n+1}) - (1 - s_{n+2}) \geq (1 - \epsilon - \mu(1 + \epsilon))c\mu^n \geq C(1 - s_n)$$

for some positive constant C. In particular, when $\sigma^2 < \infty$, there exists some constant c such that

$$\sum_{k=1}^{\infty} kp_k\left(\frac{s_{n+1}^k - s_n^k}{s_{n+2} - s_{n+1}}\right) \leq c\sum_{k=1}^{\infty} kp_k\left(\frac{1 - s_n^k}{1 - s_n}\right) \leq c\sigma^2$$

where the final inequality comes from that $(1 - s^k)(1 - s)^{-1}$ is increasing in s and converges to k for any $k \geq 1$. It therefore follows that $\mathbf{E}[\Lambda_n] \leq C(\beta\mu)^n$ so indeed

$$\mathbb{E}[\mathcal{R}_\infty] \leq C\sum_{n=0}^{\infty} \beta^{-n}(\beta\mu)^n < \infty.$$

When ξ has infinite variance but belongs to the domain of attraction of a stable law

$$\sum_{k=1}^{\infty} k p_k \left(s_{n+1}^k - s_n^k \right) = \mu \left(\left(1 - \frac{s_n f'(s_n)}{\mu} \right) - \left(1 - \frac{s_{n+1} f'(s_{n+1})}{\mu} \right) \right)$$

hence by (3.9) as $n \to \infty$ we have that $\mathbf{E}[\psi_{n+1}] \sim c\mu^{n(\alpha-2)} L_2(\mu^n)$. Combining this with (7.3) and (7.4) we have

$$\mathbf{E}[\Lambda_n] \leq C(\beta\mu)^n \mu^{n(\alpha-2)} L_2(\mu^n) = C(\beta\mu^{\alpha-1})^n L_2(\mu^n) \qquad (7.5)$$

therefore using (7.2) for C chosen sufficiently large we have that

$$\mathbb{E}[\mathcal{R}_\infty] \leq C \left(1 + \sum_{n=1}^{\infty} \mu^{n(\alpha-1)} L_2(\mu^n) \right) < \infty.$$

$\square$

We therefore have that the expected time taken for a walk started from the deepest point in a trap (of height H) to return to the deepest point is bounded above by $\mathbb{E}[\mathcal{R}_\infty] < \infty$ independently of its height. Recall that τ_x^+ is the first return time to x. The following lemma gives the probabilities of reaching the deepest point in a trap, escaping the trap from the deepest point and the transition probabilities for the walk in the trap conditional on reaching the deepest point before escaping. The proof is straightforward by comparison with the biased walk on $\mathbb{Z}$ with nearest neighbour edges so we omit it.

Lemma 7.2 *For any tree T of height $H + 1$ (with $H \geq 1$), root ρ and deepest vertex δ we have that*

$$P_{\delta_H}^T \left(\tau_\delta^+ < \tau_\rho^+ \right) = \frac{1 - \beta^{-1}}{1 - \beta^{-(H+1)}}$$

is the probability of reaching the deepest point without escaping and

$$P_\delta^T \left(\tau_\rho^+ < \tau_\delta^+ \right) = \frac{1 - \beta^{-1}}{\beta^H - \beta^{-1}}$$

is the probability of escaping from the deepest point before returning. Moreover,

$$P_{\delta_k}^T \left(\tau_{\delta_{k-1}}^+ < \tau_{\delta_{k+1}}^+ | \tau_\delta^+ < \tau_\rho^+ \right) = \frac{1 - \beta^{-(H+2-k)}}{1 - \beta^{-(H+1-k)}} \cdot \frac{\beta}{\beta + 1}$$

is the probability that the walk restricted to the spine conditioned on reaching δ before returning to ρ moves towards δ.

Since the first two probabilities are independent of the structure of the tree except for the height we write

$$p_1(H) := \frac{1 - \beta^{-1}}{1 - \beta^{-(H+1)}} \tag{7.6}$$

to be the probability that the walk reaches the deepest vertex in the tree before returning to the root starting from the bud and

$$p_2(H) := \frac{1 - \beta^{-1}}{\beta^H - \beta^{-1}} \tag{7.7}$$

to be the probability of escaping from the tree.

For the remainder of the section we will consider only the case that the offspring distribution belongs to the domain of attraction of some stable law of index $\alpha \in (1, 2)$. The first aim is to prove Proposition 7.3 which shows that the time on excursions in deep traps essentially consists of some geometric number of excursions from the deepest point to itself. We will then conclude with Corollary 7.4 which is an adaptation for FVIE and of which we omit the proof.

Recall that ρ_i^+ is the root of the ith large branch and $\tilde{\chi}_n^i$ is the time spent in this branch by the ith walk $X_n^{(i)}$. This branch has some number N^i buds which are roots of large traps where, by Proposition 3.4, N^i converges to a heavy tailed distribution. Let $\rho_{i,j}^+$ be the bud of the jth large trap $T_{i,j}^+$ in this branch then $W^{i,j} := |\{m \geq 0 : X_{m-1}^{(i)} = \rho_i^+, X_m^{(i)} = \rho_{i,j}^+\}|$ is the number of times that the jth large trap in the ith large branch is visited by the ith copy of the walk. Let $\omega^{(i,j,0)} := 0$ then for $k \leq W^{i,j}$ write $\omega^{(i,j,k)} := \min\{m > \omega^{(i,j,k-1)} : X_{m-1}^{(i)} = \rho_i^+, X_m^{(i)} = \rho_{i,j}^+\}$ to be the start time of the kth excursion into $T_{i,j}^+$ and $T^{(i,j,k)} := |\{m \in [\omega^{(i,j,k)}, \omega^{(i,j,k+1)}) : X_m^{(i)} \in T_{i,j}^+\}|$ its duration. We can then write the time spent in large traps of the ith large branch as

$$\tilde{\chi}_n^i = \sum_{j=1}^{N^i} \sum_{k=1}^{W^{i,j}} T^{(i,j,k)}.$$

For $0 \leq k \leq \mathcal{H}(T_{i,j}^+)$ write $\delta_k^{(i,j)}$ to be the spinal vertex of distance k from the deepest point in $T_{i,j}^+$. Let $T^{*(i,j,k)} := 0$ if there does not exist $m \in [\omega^{(i,j,k)}, \omega^{(i,j,k+1)}]$ such that $X_m = \delta_0^{(i,j)} =: \delta^{(i,j)}$ and

$$T^{*(i,j,k)} := \sup \left\{ m \in \left[\omega^{(i,j,k)}, \omega^{(i,j,k+1)} \right] : X_m^{(i)} = \delta^{(i,j)} \right\}$$
$$- \inf \left\{ m \in \left[\omega^{(i,j,k)}, \omega^{(i,j,k+1)} \right] : X_m^{(i)} = \delta^{(i,j)} \right\} \tag{7.8}$$

otherwise to be the duration of the kth excursion into $T_{i,j}^+$ without the first passage to the deepest point and the final passage from the deepest point to the exit. We can then define

$$\tilde{\chi}_n^{i*} := \sum_{j=1}^{N^i} \sum_{k=1}^{W^{i,j}} T^{*(i,j,k)} \tag{7.9}$$

to be the time spent in the ith large trap without the first passage to and last passage from $\delta^{(i,j)}$ on each excursion. We want to show that the difference between this and $\tilde{\chi}_n^i$ is negligible. In particular, recalling that $\mathcal{D}_n^{(n)}$ is the collection of large branches by level n, we will show that for all $t > 0$ as $n \to \infty$

$$\mathbb{P}\left(\left| \sum_{i=1}^{\left| \mathcal{D}_n^{(n)} \right|} \left(\tilde{\chi}_n^i - \tilde{\chi}_n^{i*} \right) \right| \geq t a_n^{\frac{1}{\gamma}} \right) \to 0.$$

For $\epsilon > 0$ denote

$$A_6(n) := \bigcap_{i=0}^n \left\{ \mathcal{H}\left(\mathcal{T}_{\rho_i}^{*-} \right) \leq h_{n,\epsilon}^+ \right\} \tag{7.10}$$

to be the event that there are no $h_{n,\epsilon}^+$-branches by level n. Using a union bound and (3.12) we have that $\mathbf{P}(A_6(n)^c) \leq n\mathbf{P}(\mathcal{H}(\mathcal{T}_{\rho_0}^{*-}) > h_{n,\epsilon}^+) \to 0$ as $n \to \infty$.

Write

$$A_7(n) := \bigcap_{i=0}^{\left| \mathcal{D}_n^{(n)} \right|} \left\{ N^i \leq n^{\frac{2\varepsilon}{\alpha-1}} \right\} \tag{7.11}$$

to be the event that all large branches up to level n of the backbone have fewer than $n^{\frac{2\varepsilon}{\alpha-1}}$ large traps. Conditional on the number of buds, the number of large traps in the branch follows a binomial distribution therefore

$$\mathbf{P}\left(N^i \geq n^{\frac{2\varepsilon}{\alpha-1}} \right) \leq \frac{\mathbf{P}\left(\xi^* \geq n^{\frac{1+\varepsilon/2}{\alpha-1}} \right)}{\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) > h_{n,\varepsilon} \right)}$$
$$+ \frac{\mathbf{P}\left(Bin\left(n^{\frac{1+\varepsilon/2}{\alpha-1}}, \mathbf{P}\left(\mathcal{H}\left(\mathcal{T}^\circ \right) \geq h_{n,\varepsilon} \right) \right) \geq n^{\frac{2\varepsilon}{\alpha-1}} \right)}{\mathbf{P}\left(\mathcal{H}\left(\mathcal{T}^{*-} \right) > h_{n,\varepsilon} \right)}.$$

By (4.1) $\mathbf{P}(\mathcal{H}(\mathcal{T}^{*-}) \geq h_{n,\varepsilon}) \geq Cn^{-(1-\varepsilon)}$ for n large and some constant C hence by (3.11) the first term decays faster than $n^{-\varepsilon}$. Using a Chernoff bound the second term has a stretched exponential decay. Therefore, by Lemma 4.1 and a union bound, as $n \to \infty$

$$\mathbf{P}(A_7(n)^c) \leq o(1) + Cn^\varepsilon \mathbf{P}\left(N^i \geq n^{\frac{2\varepsilon}{\alpha-1}} \right) \to 0.$$

Recall that $d_x := |c(x)|$ is the number of children of x in the tree and define

$$A_8(n) := \bigcap_{i=1}^{\left|\mathcal{D}_n^{(n)}\right|} \bigcap_{j=1}^{N^i} \left\{ \sum_{k=0}^{\mathcal{H}(\mathcal{T}_{i,j}^+)} d_{\delta_k^{(i,j)}} \leq n^{3\varepsilon/(\alpha-1)^2} \right\}$$

to be the event that there are fewer than $n^{3\varepsilon/(\alpha-1)^2}$ subtraps on the spine in any large trap. For Z_n the generation sizes associated to GW-tree $\mathcal{T}^\circ$ we have that $\mathbf{P}(Z_1 \geq n | \mathcal{H}(\mathcal{T}^\circ) \geq m)$ is non-decreasing in m; therefore, the number of offspring from a vertex on the spine of a trap can be stochastically dominated by the size biased distribution. Using this and Lemma 4.1 with the bounds on A_6 and A_7 we then have that for some slowly varying function $\overline{L}$

$$\mathbf{P}\left(A_8(n)^c\right) \leq o(1) + Cn^\varepsilon n^{\frac{2\varepsilon}{\alpha-1}} \mathbf{P}\left(\sum_{k=0}^{h_{n,\epsilon}^+} \xi_k^* \geq n^{3\varepsilon/(\alpha-1)^2}\right)$$

$$\leq o(1) + Cn^\varepsilon n^{\frac{2\varepsilon}{\alpha-1}} h_{n,\epsilon}^+ \mathbf{P}\left(\xi^* \geq n^{3\varepsilon/(\alpha-1)^2}/h_{n,\epsilon}^+\right)$$

$$\leq o(1) + n^\varepsilon n^{-\frac{\varepsilon}{\alpha-1}} \overline{L}(n)$$

where $(\xi_k^*)_{k\geq1}$ are independent variables with the size biased law; thus $\mathbf{P}(A_8(n)^c) \to 0$ as $n \to \infty$.

Proposition 7.3 *In IVIE, for any $t > 0$ as $n \to \infty$*

$$\mathbb{P}\left(\left|\sum_{i=1}^{\left|\mathcal{D}_n^{(n)}\right|} \left(\tilde{\chi}_n^i - \tilde{\chi}_n^{i*}\right)\right| \geq ta_n^{\frac{1}{\gamma}}\right) \to 0.$$

Proof Let $A'(n) := \bigcap_{i=1}^8 A_i(n)$ then using the bounds on A_i for $i = 1, \dots, 8$ it follows that $\mathbb{P}(A'(n)^c) \to 0$ as $n \to \infty$. In particular, on $A_1(n)$ (from (4.2)) we have that $|\mathcal{D}_n^{(n)}| \leq Cn^\varepsilon$ and on $A_7(n)$ (from (7.11)) we have that $N^i \leq n^{\frac{2\varepsilon}{\alpha-1}}$ for all i therefore by Markov's inequality

$$\mathbb{P}\left(\left|\sum_{i=1}^{\left|\mathcal{D}_n^{(n)}\right|} \left(\tilde{\chi}_n^i - \tilde{\chi}_n^{i*}\right)\right| \geq ta_n^{\frac{1}{\gamma}}\right)$$

$$\leq \mathbb{P}(A'(n)^c) + \frac{1}{ta_n^{\frac{1}{\gamma}}} \mathbb{E}\left[\mathbf{1}_{A'(n)} \sum_{i=1}^{\left|\mathcal{D}_n^{(n)}\right|} \sum_{j=1}^{N^i} \sum_{k=1}^{W^{i,j}} \left(T_n^{(i,j,k)} - T_n^{*(i,j,k)}\right)\right]$$

$$\leq o(1) + \frac{Cn^{\varepsilon\left(\frac{\alpha+1}{\alpha-1}\right)}}{ta_n^{\frac{1}{\gamma}}} \mathbb{E}\left[\mathbf{1}_{A'(n)} \sum_{k=1}^{W^{(1,1)}} \left(T_n^{(1,1,k)} - T_n^{*(1,1,k)}\right)\right] \qquad (7.12)$$

where we recall that $T_n^{*(i,j,k)} \leq T_n^{(i,j,k)}$ for all i, j, k.

Since the number of excursions $W^{i,j}$ are independent of the excursion times and have marginal distributions of geometric random variables with parameter $(\beta - 1)/(2\beta - 1)$

$$\mathbb{E}\left[\mathbf{1}_{A'(n)}\sum_{k=1}^{W^{(1,1)}}\left(T_n^{(1,1,k)} - T_n^{*(1,1,k)}\right)\right] = \mathbb{E}\left[W^{(1,1)}\right]\mathbb{E}\left[\mathbf{1}_{A'(n)}\left(T_n^{(1,1,1)} - T_n^{*(1,1,1)}\right)\right].$$

For a given excursion either the walk reaches the deepest point $\delta^{(1,1)}$ before returning to the root $\rho_{1,1}^+$ or it doesn't. In the first case the difference $T_n^{(1,1,1)} - T_n^{*(1,1,1)}$ is the time taken to reach $\delta^{(1,1)}$ conditional on the walker reaching $\delta^{(1,1)}$ before $\rho_{1,1}^+$ added to the time taken to reach $\rho_{1,1}^+$ from $\delta^{(1,1)}$ conditional on reaching $\rho_{1,1}^+$ before returning to $\delta^{(1,1)}$. In the second case the difference is the time taken to return to the root given that the walker returns to the root without reaching $\delta^{(1,1)}$. In particular, recalling that $\mathcal{T}_{1,1}^+$ is the trap rooted at $\rho_{1,1}^+$ we have that

$$\mathbb{E}[\mathbf{1}_{A'(n)}(T_n^{(1,1,1)} - T_n^{*(1,1,1)})] \leq \mathbf{E}\left[\mathbf{1}_{A'(n)}E_{\rho_{1,1}^+}^{\mathcal{T}_{1,1}^+}\left[\mathbf{1}_{A'(n)}\tau_{\delta^{(1,1)}}^+\big|\tau_{\delta^{(1,1)}}^+ < \tau_{\rho_{1,1}^+}^+\right]\right]$$

$$+ \mathbf{E}\left[\mathbf{1}_{A'(n)}E_{\delta^{(1,1)}}^{\mathcal{T}_{1,1}^+}\left[\mathbf{1}_{A'(n)}\tau_{\rho_{1,1}^+}^+\big|\tau_{\rho_{1,1}^+}^+ < \tau_{\delta^{(1,1)}}^+\right]\right]$$

$$+ \mathbf{E}\left[\mathbf{1}_{A'(n)}E_{\rho_{1,1}^+}^{\mathcal{T}_{1,1}^+}\left[\mathbf{1}_{A'(n)}\tau_{\rho_{1,1}^+}^+\big|\tau_{\rho_{1,1}^+}^+ < \tau_{\delta^{(1,1)}}^+\right]\right].$$

$$(7.13)$$

We want to show that each of the terms in (7.13) can be bounded appropriately. This follows similarly to Lemmas 8.2 and 8.3 of [4] so we only sketch the details. Conditional on the event that the walk returns to the root of the trap before reaching the deepest point we have that:

1. the transition probabilities of the walk in subtraps are unchanged,
2. from any vertex on the spine, the walk is more likely to move towards the root than to any vertex in the subtrap,
3. from any vertex on the spine, excluding the root and deepest point, the probability of moving towards the root is at least β times that of moving towards the deepest point.

Property 3 above shows that the probability of escaping the trap from any vertex on the spine is at least the probability p_∞ of a regeneration for the β-biased random walk on $\mathbb{Z}$. From this we have that the number of visits to any spinal vertex can be stochastically dominated by a geometric random variable with parameter p_∞. Similarly, using property 2 above, we see that the number of visits to any subtrap can be stochastically dominated by a geometric random variable with parameter $p_\infty/2$.

Using a union bound with A_1, A_7, A_8 and (3.1) we have that with high probability there are no subtraps of height greater than $h_{n,\varepsilon}$. In particular, by (5.2), the expected time in any subtrap can be bounded above by $C(\beta\mu)^{h_{n,\varepsilon}}$ for some constant C using property 1. From this it follows that

$$\mathbf{E}\left[\mathbf{1}_{A'(n)}E_{\rho_{1,1}^+}^{\mathcal{T}_{1,1}^+}\left[\mathbf{1}_{A'(n)}\tau_{\rho_{1,1}^+}^+\,\big|\,\tau_{\rho_{1,1}^+}^+ < \tau_{\delta(1,1)}^+\right]\right]$$

$$\leq \mathbf{E}\left[\mathbf{1}_{A'(n)}E_{\delta(1,1)}^{\mathcal{T}_{1,1}^+}\left[\mathbf{1}_{A'(n)}\tau_{\rho_{1,1}^+}^+\,\big|\,\tau_{\rho_{1,1}^+}^+ < \tau_{\delta(1,1)}^+\right]\right]$$

$$\leq o(1) + h_{n,\varepsilon}^+ E\left[Geo(p_\infty)\right] + Cn^{\frac{3\varepsilon}{(\alpha-1)^2}}(\beta\mu)^{h_{n,\varepsilon}}$$

$$\leq o(1) + C\overline{L}(n)n^{\frac{(1-\varepsilon)}{\alpha-1}\frac{\log(\beta\mu)}{\log(\mu^{-1})}+\frac{3\varepsilon}{(\alpha-1)^2}}$$

for some constant C and slowly varying function $\overline{L}$.

A symmetric argument shows that the same bound can be achieved for the first term in (7.13). It then follows that the second term in (7.12) can be bounded above by $C_t L_1(n)n^{-\frac{1}{\alpha-1}+\tilde{\varepsilon}}$ where $\tilde{\varepsilon}$ can be made arbitrarily small by choosing ε sufficiently small. □

A straightforward adaptation of Proposition 8.1 of [4] (similar to the previous calculation) shows Corollary 7.4 which is the corresponding result for FVIE.

Corollary 7.4 *In FVIE, for any $t > 0$ as $n \to \infty$*

$$\mathbb{P}\left(\left|\sum_{i=1}^{\left|\mathcal{D}_n^{(n)}\right|}\left(\tilde{\chi}_n^i - \tilde{\chi}_n^{i*}\right)\right| \geq tn^{\frac{1}{\gamma}}\right) \to 0.$$

By Proposition 7.3 and Corollary 7.4, in FVIE and IVIE, almost all time up to the walk reaching level n is spent on excursions from the deepest point in deep traps. The aim of the remainder of the section is to prove Proposition 7.14 which shows that the time spent on the excursions from the deepest point in a single large branch (suitably scaled) converges in distribution along the given subsequences. To ease notation, for the remainder of the section we work on a dummy branch $\mathcal{T}^*$ so that the time $\tilde{\chi}_n^{i*}$ has the distribution of a sum of excursion times from the deepest points of $\mathcal{T}^*$.

Recall from Definition 4 that $\mathcal{T}^{*-}$ is a dummy branch with root ρ, buds $\rho_1, \ldots, \rho_{\xi^*-1}$ each of which is the root of an f-GW-tree $\mathcal{T}_j^\circ$ with height $H_j := \mathcal{H}(\mathcal{T}_j^\circ)$. We now define a pruned version of this branch which only contains traps of height at least $h_{n,\varepsilon}$.

Definition 8 (*Pruned dummy branch*) Let

$$N := \sum_{j=1}^{\xi^*-1}\mathbf{1}_{\{H_j\geq h_{n,\varepsilon}\}} \tag{7.14}$$

denote the number of traps in $\mathcal{T}^{*-}$ of at least critical height. Denote $(\mathcal{T}_j^+)_{j=1}^N$ to be those large traps, $(\rho_j^+)_{j=1}^N$ their roots and $H_j^+ := \mathcal{H}(\mathcal{T}_j^+)$ the height of the jth large trap in the branch. Similarly, let $(\mathcal{T}_j^-)_{j=1}^{\xi^*-1-N}$ denote the small traps, $(\rho_j^-)_{j=1}^{\xi^*-1-N}$ their roots and $H_j^- := \mathcal{H}(\mathcal{T}_j^-)$ the height of the jth small trap in the branch.

Let $\mathcal{T}^*$ be $\mathcal{T}^{*-}$ pruned to consist precisely of the root ρ, buds $(\rho_j^+)_{j=1}^N$ and traps $(\mathcal{T}_j^+)_{j=1}^N$.

We write $\overline{H} := \mathcal{H}(\mathcal{T}^*) - 1$ to be the height of the largest trap and for $K \in \mathbb{Z}$ let $\overline{H}_n^K := h_{n,0} + K$ then denote

$$\mathbb{P}^K(\cdot) := \mathbb{P}\left(\cdot \middle| \overline{H} = \overline{H}_n^K\right) \quad \text{and} \quad \mathbf{P}^K(\cdot) := \mathbf{P}\left(\cdot \middle| \overline{H} = \overline{H}_n^K\right).$$

Write W^j to be the total number of excursions into $\mathcal{T}_j^+$ and B^j the number of excursions which reach the deepest point δ^j.

For each $k \leq B^j$ we define $G^{j,k}$ to be the number of return times to δ^j on the kth excursion which reaches δ^j.

For $l = 1, \ldots, G^{j,k}$ let $\mathcal{R}^{j,k,l}$ denote the duration of the lth excursion from δ^j to itself on the kth excursion into $\mathcal{T}_j^+$ which reaches δ^j.

The height of the branch and the total number of traps in the branch have a strong relationship. Lemma 7.5 shows the exact form of this relationship in the limit as $n \to \infty$. Recall from (3.1) that c_μ is the positive constant such that $\mathbf{P}(\mathcal{H}(\mathcal{T}) \geq n) \sim c_\mu \mu^n$ as $n \to \infty$ then write

$$b_n^K := \frac{\mu^{-\overline{H}_n^K}}{c_\mu}. \tag{7.15}$$

Lemma 7.5 *In IVIE, under $\mathbf{P}^K$ we have that the sequence of random variables $(\xi^* - 1)/b_n^K$ converge in distribution to some random variable $\overline{\xi}$ satisfying*

$$\mathbf{P}\left(\overline{\xi} \geq t\right) = \frac{\alpha - 1}{\Gamma(2 - \alpha)\left(1 - \mu^{\alpha-1}\right)} \int_t^\infty y^{-\alpha} \left(e^{-\mu y} - e^{-y}\right) \mathrm{d}y.$$

Proof We prove this by showing the convergence of

$$\mathbf{P}\left(\xi^* - 1 \geq t b_n^K \middle| \overline{H} = \overline{H}_n^K\right) = \mathbf{P}\left(\overline{H} = \overline{H}_n^K \middle| \xi^* - 1 \geq t b_n^K\right) \frac{\mathbf{P}\left(\xi^* - 1 \geq t b_n^K\right)}{\mathbf{P}\left(\overline{H} = \overline{H}_n^K\right)} \tag{7.16}$$

for all $t > 0$. To begin we consider $\mathbf{P}\left(\overline{H} = \overline{H}_n^K \middle| \xi^* - 1 \geq t b_n^K\right)$.

The heights of individual traps are independent under this conditioning hence

$$\mathbf{P}\left(\overline{H} \leq \overline{H}_n^K \middle| \xi^* - 1 \geq t b_n^K\right) = \mathbf{E}\left[\mathbf{P}\left(\mathcal{H}(\mathcal{T}^\circ) \leq \overline{H}_n^K\right)^{\xi^* - 1} \middle| \xi^* - 1 \geq t b_n^K\right].$$

We know the asymptotic form of $\mathbf{P}(\mathcal{H}(\mathcal{T}^\circ) \leq \overline{H}_n^K)$ from (3.1) thus we need to consider the distribution of $\xi^* - 1$ conditioned on $\xi^* - 1 \geq t b_n^K$. By the tail formula for $\xi^* - 1$ following Definition 3 we have that for $r \geq 1$ as $n \to \infty$

$$\mathbf{P}\left(\frac{\xi^* - 1}{tb_n^K} \geq r \Big| \xi^* - 1 \geq tb_n^K\right) = \frac{\mathbf{P}\left(\xi^* - 1 \geq rtb_n^K\right)}{\mathbf{P}\left(\xi^* - 1 \geq tb_n^K\right)} \sim r^{-(\alpha - 1)}.$$

We therefore have that, conditional on $\xi^* - 1 \geq tb_n^K$, the sequence $(\xi^* - 1)/tb_n^K$ converges in distribution to a variable Y with tail $\mathbf{P}(Y \geq r) = r^{-(\alpha - 1)} \wedge 1$. Using the form of b_n^K we then have that

$$\mathbf{P}\left(\mathcal{H}(\mathcal{T}^\circ) \leq \overline{H}_n^K\right)^{tb_n^K} = e^{-t\mu(1 + o(1))}.$$

It therefore follows that

$$\lim_{n \to \infty} \mathbf{P}\left(\overline{H} \leq \overline{H}_n^K | \xi^* - 1 \geq tb_n^K\right) = \mathbf{E}\left[e^{-t\mu Y}\right].$$

Repeating with $\overline{H}_n^K$ replaced by $\overline{H}_n^K - 1$ we have that $\mathbf{P}(\overline{H} = \overline{H}_n^K | \xi^* - 1 \geq tb_n^K) \to \mathbf{E}[e^{-t\mu Y}] - \mathbf{E}[e^{-tY}]$ as $n \to \infty$. For $\theta > 0$

$$\mathbf{E}\left[e^{-\theta tY}\right] = (\alpha - 1)t^{\alpha - 1} \int_t^\infty e^{-\theta y} y^{-\alpha} dy$$

therefore

$$\lim_{n \to \infty} \mathbf{P}\left(\overline{H} = \overline{H}_n^K | \xi^* - 1 \geq tb_n^K\right) = (\alpha - 1)t^{\alpha - 1} \int_t^\infty y^{-\alpha}(e^{-\mu y} - e^{-y}) dy. \tag{7.17}$$

By (3.12) we have that as $n \to \infty$

$$\mathbf{P}\left(\overline{H} = \overline{H}_n^K\right) = \mathbf{P}\left(\mathcal{H}\left(\mathcal{T}^{*-}\right) > \overline{H}_n^K\right) - \mathbf{P}\left(\mathcal{H}\left(\mathcal{T}^{*-}\right) > \overline{H}_n^K + 1\right)$$

$$\sim \Gamma(2 - \alpha) c_\mu^{\alpha - 1}\left(1 - \mu^{\alpha - 1}\right) \mathbf{P}\left(\xi^* - 1 \geq \mu^{-\overline{H}_n^K}\right)$$

$$= \Gamma(2 - \alpha) c_\mu^{\alpha - 1}\left(1 - \mu^{\alpha - 1}\right) \mathbf{P}\left(\xi^* - 1 \geq c_\mu b_n^K\right)$$

therefore

$$\frac{\mathbf{P}\left(\xi^* - 1 \geq tb_n^K\right)}{\mathbf{P}\left(\overline{H} = \overline{H}_n^K\right)} \sim \frac{\mathbf{P}\left(\xi^* - 1 \geq tb_n^K\right)}{\Gamma(2 - \alpha)\left(1 - \mu^{\alpha - 1}\right) c_\mu^{\alpha - 1} \mathbf{P}\left(\xi^* - 1 \geq c_\mu b_n^K\right)}$$

$$\sim \frac{t^{-(\alpha - 1)}}{\Gamma(2 - \alpha)\left(1 - \mu^{\alpha - 1}\right)}.$$

Combining this with (7.17) in (7.16) we have that

$$\lim_{n \to \infty} \mathbf{P}\left(\xi^* - 1 \geq tb_n^K | \overline{H} = \overline{H}_n^K\right) = \frac{\alpha - 1}{\Gamma(2 - \alpha)\left(1 - \mu^{\alpha - 1}\right)} \int_t^\infty y^{-\alpha}\left(e^{-\mu y} - e^{-y}\right) dy.$$

$$\square$$

Notice that under $\mathbf{P}$ the pruned dummy branch $\mathcal{T}^*$ is the single vertex ρ with high probability however under $\mathbf{P}^K$ there is at least one trap. By Lemma 5.1, conditional on N, $(W^j)_{j=1}^N$ have a joint negative multinomial distribution. Moreover, W_j, B_j are coupled so that B^j is binomially distributed with W^j trials and success probability $p_1(H_j^+)$. The number $G^{j,k}$ of return times to δ^j is geometrically distributed with failure probability $p_2(H_j^+)$. It follows that each $\tilde{\chi}_n^{i*}$ is equal in distribution to

$$\chi_n^* := \sum_{j=1}^{N} \sum_{k=1}^{B^j} \sum_{l=1}^{G^{j,k}} \mathcal{R}^{j,k,l}.$$

Define the scaled excursion time in large traps of a large branch as

$$\zeta^{(n)} = \chi_n^* \beta^{-\overline{H}} = \beta^{-\overline{H}} \sum_{j=1}^{N} \sum_{k=1}^{B^j} \sum_{l=1}^{G^{j,k}} \mathcal{R}^{j,k,l} \qquad (7.18)$$

then we will show that $\zeta^{(n)}$ converges in distribution under $\mathbb{P}^K$ along subsequences $n_l(t)$. Lemma 7.6 gives an upper bound on the number of large traps in a branch conditioned on its height.

Lemma 7.6 *For any $\epsilon > 0$ and $K \in \mathbb{Z}$*

$$\lim_{n \to \infty} \mathbf{P}^K \left(N \geq n^{\frac{\varepsilon + \epsilon}{\alpha - 1}} \right) = 0.$$

Proof Conditioned on the height of the branch and number of buds we have that at least one trap attains the maximum height, all others have the distribution of heights of GW-tree conditioned on their maximum height therefore

$$\mathbf{P}^K \left(N \geq n^{\frac{\varepsilon + \epsilon}{\alpha - 1}} \right) \leq \mathbf{P}^K \left(\xi^* - 1 \geq \log(n) b_n^K \right)$$
$$+ \mathbf{P} \left(N \geq n^{\frac{\varepsilon + \epsilon}{\alpha - 1}} - 1 \Big| \xi^* - 1 = \log(n) b_n^K \right). \qquad (7.19)$$

By Lemma 7.5 $\mathbf{P}^K \left(\xi^* - 1 \geq \log(n) b_n^K \right)$ converges to 0 as $n \to \infty$. Conditioned on having $\xi^* - 1 = \log(n) b_n^K$ buds we have that N is binomially distributed with $\log(n) b_n^K$ trails and success probability $\mathbf{P}(\mathcal{H}(\mathcal{T}^\circ) \geq h_{n,\varepsilon}) \leq C \mu^{h_{n,\varepsilon}}$ by (3.1). Since for some slowly varying function $\overline{L}$ we have that

$$\mathbf{E} \left[Bin \left(\log(n) b_n^K, \ C\mu^{h_{n,\varepsilon}} \right) \right] \leq C\mu^K \log(n) \frac{a_n}{a_{n^{1-\varepsilon}}} \leq \overline{L}(n) \mu^K n^{\frac{\varepsilon}{\alpha - 1}},$$

a Chernoff bound shows that the final term in (7.19) converges to 0. $\square$

For $\tilde{\varepsilon} > 0$ write

$$A_9(n) = \bigcap_{j=1}^{N} \left\{ 1 \leq \frac{\beta^{H_j^+}}{1 - \beta^{-1}} E\left[G^{j,1}\right]^{-1} \leq 1 + \tilde{\varepsilon} \right\}.$$

Recall from (7.7) that $p_2(H)$ is the probability that a walk started from the deepest point of a tree of height H reaches the root before returning to the deepest point. Since $G^{j,k}$ are independent geometric random variables there exist independent exponential random variables $e_{j,k}$ such that

$$G^{j,k} = \left\lfloor \frac{e_{j,k}}{-\log\left(1 - p_2(H_j^+)\right)} \right\rfloor \sim Geo\left(p_2(H_j^+)\right).$$

By (7.7) we then have that

$$E[G^{j,1}] = \left(1 - \frac{1 - \beta^{-1}}{\beta^{H_j^+} - \beta^{-1}}\right) \left(1 - \beta^{-\left(H_j^+ + 1\right)}\right) \frac{\beta^{H_j^+}}{1 - \beta^{-1}} \tag{7.20}$$

therefore, since $H_j^+ \geq h_{n,\varepsilon}$, for any $\tilde{\varepsilon} > 0$ there exists n large such that $\mathbf{P}^K(A_9(n)) = 1$ for any $K \in \mathbb{Z}$.

Recall from (7.7) and Definition 8 that $G^{j,k}$ is geometrically distributed with failure probability $p_2(H_j^+) \geq p_2(h_{n,\varepsilon})$. Write

$$A_{10}^{(j,k)}(n) := \left\{ (1 - \tilde{\varepsilon})\, G^{j,k} \leq E\left[G^{j,k}\right] e_{j,k} \leq (1 + \tilde{\varepsilon})\, G^{j,k} \right\}.$$

Then, using convergence of scaled geometric variables to exponential variables (see the proof of part (3) of Proposition 9.1 in [4]), we have that there exists a constant $\tilde{C}$ such that for any $\tilde{\varepsilon} > 0$ there exists n large such that

$$P\left(A_{10}^{(j,k)}(n)^c\right) \leq \tilde{C} p_2(h_{n,\varepsilon}) \leq \tilde{C} a_{n^{1-\varepsilon}}^{-1/\gamma}.$$

By Definition 8 we have that $B_j \leq W_j$. Moreover $N \leq n^{\frac{\varepsilon + \tilde{\varepsilon}}{\alpha - 1}}$ with high probability for any $\tilde{\varepsilon} > 0$ by Lemma 7.6 and $W^j \leq C \log(n)$ for all j by the bound on the event $A_5(n)^c$ (from (5.3)). Therefore, writing

$$A_{10}(n) := \bigcap_{j=1}^{N} \bigcap_{k=1}^{B^j} A_{10}^{(j,k)}(n)$$

a union bound gives us that $P(A_{10}(n)^c) \to 0$ as $n \to \infty$.

By comparison with the biased random walk on $\mathbb{Z}$ we have that $p_1(H_j^+) \geq p_\infty = 1 - \beta^{-1}$ therefore we can define a random variable $B_\infty^j \sim Bin(B^j, p_\infty / p_1(H_j^+))$. It then follows that $B^j \geq B_\infty^j \sim Bin(W^j, p_\infty)$ and

$$p_1(H_j^+) - p_\infty = \frac{1 - \beta^{-1}}{1 - \beta^{-(H_j^+ + 1)}} - (1 - \beta^{-1}) \leq \beta^{-H_j^+}. \tag{7.21}$$

Write

$$A_{11}(n) := \bigcap_{j=1}^{N} \left\{ B^j = B_\infty^j \right\}.$$

Since the marginal distribution of W^1 doesn't depend on n, using (7.21), the bound on N from Lemma 7.6 and the coupling between B^1 and B_∞^1 we have that

$$\mathbb{P}^K(A_{11}(n)^c) \leq o(1) + n^{\frac{\varepsilon + \bar{\varepsilon}}{\alpha - 1}} \sum_{k=0}^{\infty} \mathbb{P}(W^1 = k) \mathbb{P}(B^1 \neq B_\infty^1 | W^1 = k)$$

$$\leq o(1) + n^{\frac{\varepsilon + \bar{\varepsilon}}{\alpha - 1}} \sum_{k=0}^{\infty} \mathbb{P}(W^1 = k) k \left(p_1(H_1^+) - p_\infty \right)$$

$$\leq o(1) + n^{\frac{\varepsilon + \bar{\varepsilon}}{\alpha - 1}} \beta^{-h_{n,\varepsilon}} \mathbb{E}\left[W^1 \right] \tag{7.22}$$

which decays to 0 as $n \to \infty$.

By choosing $\varepsilon > 0$ sufficiently small we can choose κ in the range $\varepsilon(1/\gamma + 1/(\alpha - 1)) < \kappa < \min\{2(\alpha - 1), \ 1/\gamma\}$ then write

$$A_{12}(n) := \bigcap_{j=1}^{N} \left\{ E\left[(\mathcal{R}_n^{j,1,1})^2 \right] < n^{\frac{\gamma^{-1} - \kappa}{\alpha - 1}} \right\}$$

to be the event that there are no large traps with expected squared excursion time too large.

Lemma 7.7 *In IVIE, for any $K \in \mathbb{Z}$, as $n \to \infty$ we have that $\mathbb{P}^K(A_{12}(n)^c) \to 0$.*

Proof Recall from (7.10) that, for $\epsilon > 0$, $A_6(n)$ is the event that all large branches are shorter than $h_{n,\epsilon}^+$ and since $N \leq n^{\frac{\varepsilon + \bar{\varepsilon}}{\alpha - 1}}$ with high probability we have that

$$\mathbb{P}(A_{12}(n)^c) \leq o(1) + n^{\frac{\varepsilon + \bar{\varepsilon}}{\alpha - 1}} \mathbb{P}\left(\mathbf{1}_{\{A_6(n)\}} E\left[(\mathcal{R}_n^{1,1,1})^2 \right]^{1/2} > n^{\frac{\gamma^{-1} - \kappa}{2(\alpha - 1)}} \right).$$

A straightforward argument using conductances (see the proof of Lemma 9.1 in [4]) gives

$$E\left[\left(\mathcal{R}_n^{1,1,1}\right)^2\right]^{1/2} \leq C \sum_{y \in \mathcal{T}_1^+} \beta^{d(y,\delta^1)/2} \pi(y)$$

where π is the invariant measure scaled so that $\pi(\delta^1) = 1$ and d denotes the graph distance. We then have that

$$\mathbf{E}\left[\mathbf{1}_{\{A_6(n)\}} E\left[\left(\mathcal{R}_n^{(1,1,1)}\right)^2\right]^{1/2}\right] \leq C\mathbf{E}\left[\mathbf{1}_{\{A_6(n)\}} \sum_{y \in \mathcal{T}_1^+} \beta^{d(y,\delta^1)/2} \pi(y)\right]$$

$$\leq C\mathbf{E}\left[\mathbf{1}_{\{A_6(n)\}} \sum_{i \geq 1} \beta^{i/2} \beta^{-i}(1 + \Lambda_i)\right]$$

$$\leq C \sum_{i=0}^{h_{n,\epsilon}^+} \left(\beta^{1/2} \mu^{\alpha-1-\epsilon}\right)^i$$

where the final inequality follows by (7.5). If $\beta^{1/2}\mu^{\alpha-1-\epsilon} \leq 1$ then by Markov's inequality we have that $\mathbb{P}^K(A_{12}(n)^c) \to 0$ as $n \to \infty$ since $\kappa < \gamma^{-1}$. Otherwise by Markov's inequality

$$\mathbb{P}^K(A_{12}(n)^c) \leq o(1) + Cn^{\frac{\varepsilon+\tilde{\epsilon}}{\alpha-1}}\left(\beta^{1/2}\mu^{\alpha-1-\epsilon}\right)^{h_{n,\epsilon}^+} n^{\frac{\kappa-\gamma^{-1}}{2(\alpha-1)}}$$

$$\leq \overline{L}(n)n^{\frac{\kappa}{2(\alpha-1)}-1+\frac{\epsilon}{\alpha-1}\left(\frac{1}{2\gamma}+2-\alpha+\epsilon\right)+\frac{\varepsilon+\tilde{\epsilon}}{\alpha-1}}$$

for some slowly varying function $\overline{L}$. In particular, since $\kappa < 2(\alpha-1)$ we can choose $\epsilon, \varepsilon, \tilde{\epsilon}$ sufficiently small such that this converges to 0 as $n \to \infty$. $\square$

Write

$$A_{13}(n) = \bigcap_{j=1}^{N} \bigcap_{k=1}^{B^j} \left\{(1-\tilde{\varepsilon})G^{j,k}E\left[\mathcal{R}_n^{j,1,1}\right] \leq \sum_{l=1}^{G^{j,k}} \mathcal{R}^{j,k,l} \leq (1+\tilde{\varepsilon})G^{j,k}E\left[\mathcal{R}_n^{j,1,1}\right]\right\}$$

to be the event that on each excursion that reaches the deepest point of a large trap, the total excursion time before leaving the trap is approximately the product of the number of excursions and the expected excursion time.

Lemma 7.8 *In IVIE, for any $K \in \mathbb{Z}$, as $n \to \infty$ we have that $\mathbb{P}^K(A_{13}(n)^c) \to 0$.*

Proof With high probability we have that no trap is visited more than $C\log(n)$ by (5.3) and also $N \leq n^{\frac{\varepsilon+\tilde{\epsilon}}{\alpha-1}}$ by Lemma 7.6. Any excursion is of length at least 2 hence $E[\mathcal{R}_n^{1,1,1}] \geq 2$. Therefore, by Lemma 7.7 and Chebyshev's inequality

$$\mathbb{P}^K(A_{13}(n)^c) \le o(1)$$

$$+ C \log(n) n^{\frac{\varepsilon + \tilde{\varepsilon}}{\alpha - 1}} \mathbb{P}\left(\left|\sum_{l=1}^{G^{1,1}} \frac{\mathcal{R}_n^{1,1,l}}{E\left[\mathcal{R}_n^{1,1,1}\right] G^{1,1}} - 1\right| > \tilde{\varepsilon}, G^{1,1} > 0, E\left[\left(\mathcal{R}_n^{1,1,1}\right)^2\right] < n^{\frac{\gamma^{-1} - \kappa}{\alpha - 1}}\right)$$

$$\le o(1) + \frac{C \log(n) n^{\frac{\gamma^{-1} + \varepsilon + \tilde{\varepsilon} - \kappa}{\alpha - 1}}}{\tilde{\varepsilon}^2} E\left[\frac{\mathbf{1}_{\{G^{1,1} > 0\}}}{G^{1,1}}\right].$$

It then follows that since $G^{1,1} \sim Geo(p_2(H_1^+))$ (where from (7.7) $p_2(H)$ is the probability that a walk reaches the deepest point in the trap of height H) and $p_2(H_1^+) \le c\beta^{-h_{n,\varepsilon}} = ca_n^{-\frac{1}{\gamma}} {}^{1-\varepsilon}$

$$E\left[\frac{\mathbf{1}_{\{G^{(1,1,1)} > 0\}}}{G^{(1,1,1)}}\right] \le E\left[-\frac{p_2(H_1^+)}{1 - p_2(H_1^+)} \log\left(p_2(H_1^+)\right)\right] \le \overline{L}(n) n^{-\frac{1-\varepsilon}{\gamma(\alpha-1)}}$$

for some slowly varying function $\overline{L}$. In particular, $\mathbb{P}^K(A_{13}(n)^c) \le o(1) + L_{\tilde{\varepsilon}}(n) n^{\frac{\varepsilon\left(\frac{1}{\gamma} + \frac{1}{\alpha-1}\right) + \tilde{\varepsilon} - \kappa}{\alpha - 1}}$ which converges to zero by the choice of $\kappa > \varepsilon(1/\gamma + 1/(\alpha - 1))$. $\square$

Lemma 7.9 illustrates that the expected time spent on an excursion from the deepest point of a trap of height at least $h_{n,\varepsilon}$ doesn't differ too greatly from the expected excursion time in an infinite version of the trap. Let $\mathcal{R}_\infty^j$ be an excursion time from δ^j to itself in an extension of $\mathcal{T}_j^+$ to an infinite trap constructed according to the algorithm at the beginning of the section where $\mathcal{T}_{H_j^+}^{\prec}$ is replaced by $\mathcal{T}_j^+$. Write

$$A_{14}(n) := \bigcap_{j=1}^N \left\{E\left[\mathcal{R}_\infty^j\right] - E\left[\mathcal{R}^{j,k,l}\right] < \tilde{\varepsilon}\right\}.$$

Lemma 7.9 *In IVIE, for any $K \in \mathbb{Z}$ as $n \to \infty$ we have that $\mathbf{P}^K(A_{14}(n)^c) \to 0$.*

Proof A straightforward computation similar to that in Proposition 9.1 of [4] yields that for some constant c and n sufficiently large

$$0 \le E\left[\mathcal{R}_\infty^j\right] - E\left[\mathcal{R}^{j,k,l}\right] \le c\beta^{-h_{n,\varepsilon}/2} \sum_{k=0}^{h_{n,\varepsilon}/2} \beta^{-k}(1 + \Lambda_k) + 2 \sum_{k=h_{n,\varepsilon}/2+1}^{\infty} \beta^{-k}(1 + \Lambda_k)$$

for all $j = 1, \ldots, N$ where Λ_k are the weights of the extension of $\mathcal{T}_j^+$. Recall that $N \le n^{\frac{\varepsilon + \tilde{\varepsilon}}{\alpha - 1}}$ with high probability by Lemma 7.6, therefore by (7.5) and Markov's inequality

$$\mathbf{P}(A_{14}(n)^c) \leq \frac{Cn^{\frac{\varepsilon+\tilde{\varepsilon}}{\alpha-1}}}{\tilde{\varepsilon}} \mathbf{E}\left[E\left[\mathcal{R}_\infty^j\right] - E\left[\mathcal{R}_n^{j,1,1}\right]\right]$$

$$\leq C_{\tilde{\varepsilon}} n^{\frac{\varepsilon+\tilde{\varepsilon}}{\alpha-1}}\left(\beta^{-\frac{h_{n,\varepsilon}}{2}}\sum_{k=0}^{\infty}\left(\beta^{-k} + \mu^{k(\alpha-1-\tilde{\varepsilon})}\right) + \sum_{k=h_{n,\varepsilon}/2+1}^{\infty}\mu^{k(\alpha-1-\tilde{\varepsilon})}\right)$$

$$\leq C_{\tilde{\varepsilon}} n^{\frac{\varepsilon+\tilde{\varepsilon}}{\alpha-1}}\left(\beta^{-\frac{h_{n,\varepsilon}}{2}} + \mu^{h_{n,\varepsilon}\frac{(\alpha-1-\tilde{\varepsilon})}{2}}\right).$$

Since we can choose $\tilde{\varepsilon}$, ε and $\tilde{\epsilon}$ arbitrarily small we indeed have the desired result. □

Define

$$Z_\infty^n := \frac{1}{1-\beta^{-1}}\sum_{j=1}^{N}\beta^{H_j^+ - \overline{H}}E\left[\mathcal{R}_\infty^j\right]\sum_{k=1}^{B_\infty^j}e_{j,k} \tag{7.23}$$

whose distribution depends on n only through N and $(H_j^+ - \overline{H})_{j=1}^N$. Recalling the definition of $\zeta^{(n)}$ in (7.18), since $e_{j,k}$ are the exponential random variables defining $G^{j,k}$, $B_\infty^j \sim Bin(B^j, p_\infty/p_1(H_1^+))$ and the random variable N is the same in both equations, we have that $\zeta^{(n)}$ and Z_∞^n are defined on the same probability space.

Proposition 7.10 *In IVIE, for any $K \in \mathbb{Z}$ and $\tilde{\epsilon} > 0$*

$$\lim_{n\to\infty}\mathbb{P}^K\left(\left|\zeta^{(n)} - Z_\infty^n\right| > \tilde{\epsilon}\right) = 0.$$

Proof Using the bounds on A_{11}, A_{13} and A_{14} from (7.22) and Lemmas 7.8 and 7.9 respectively there exists some function $g : \mathbb{R} \to \mathbb{R}$ such that $\lim_{\tilde{\varepsilon}\to 0+} g(\tilde{\varepsilon}) = 0$ and for sufficiently large n (independently of K)

$$\mathbb{P}^K\left(\left|\zeta^{(n)} - Z_\infty^n\right| > \tilde{\epsilon}\right) \leq o(1) + 2\mathbb{P}^K\left(g(\tilde{\varepsilon})Z_\infty^n > \tilde{\epsilon}\right).$$

It therefore suffices to show that $(Z_\infty^n)_{n\geq 0}$ are tight under $\mathbb{P}^K$. Write

$$S_j := \frac{1}{1-\beta^{-1}}E\left[\mathcal{R}_\infty^j\right]\sum_{k=1}^{B_\infty^j}e_{j,k}. \tag{7.24}$$

The variables $E[\mathcal{R}_\infty^j]$, B_∞^j and $e_{j,k}$ are independent, don't depend on K and have finite mean (by Lemma 7.1, the geometric distribution of W^j and exponential distribution of $e^{j,k}$) therefore

$$\mathbb{E}^K\left[S_j\right] \leq C < \infty \tag{7.25}$$

uniformly over K. We can then write

$$Z_\infty^n = \sum_{j=1}^{N} \beta^{H_j^+ - \overline{H}_n^K} S_j.$$

The distribution of S_j is independent of the height of the trap. The number of large traps N is dominated by the total number of traps $\xi^* - 1$ in the branch thus reintroducing small traps

$$\mathbb{P}^K \left(Z_\infty^n \geq t \right) \leq \mathbb{P}^K \left(\sum_{j=1}^{b_n^K \log(t)} \beta^{H_j - \overline{H}_n^K} S_j \geq t \right) + \mathbb{P}^K \left(\xi^* - 1 \geq b_n^K \log(t) \right) \quad (7.26)$$

where we recall that, under $\mathbb{P}^K$, $(H_j)_{j=1}^{\xi^*-1}$ are distributed as the heights of independent f-GW-trees conditioned so that the largest is of height $\overline{H}_n^K$ and $(S_j)_{j=1}^{\xi^*-1}$ are i.i.d. with the law of S_1. By Lemma 7.5 we have that $\lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}^K(\xi^* - 1 \geq b_n^K \log(t)) = 0$ therefore it remains to bound the first term in (7.26).

Write $\Phi = \inf\{r \geq 1 : H_r = \overline{H}_n^K\}$ to be the index of the first trap with height the same as the maximum in the branch. Conditional on trap j being the first in the branch which attains the maximum height we have that the heights of the remaining traps are independent and either at most the height of the largest (for higher indices than j) or strictly shorter (for lower indices than j). In particular, this means that

$$\mathbb{P}^K \left(\sum_{j=1}^{b_n^K \log(t)} \beta^{H_j - \overline{H}_n^K} S_j \geq t \right) \leq \mathbb{P}^K \left(\sum_{j=1}^{b_n^K \log(t)} \beta^{H_j - \overline{H}_n^K} S_j \geq t \,\middle|\, \Phi = 1 \right)$$

$$\leq \mathbb{P} \left(S_1 \geq \log(t) \right) + \mathbb{P} \left(\sum_{j=2}^{b_n^K \log(t)} \beta^{H_j - \overline{H}_n^K} S_j \geq t - \log(t) \,\middle|\, H_j \leq \overline{H}_n^K \, \forall j \geq 2 \right).$$

The distribution of S_1 is independent of n therefore $\lim_{t\to\infty} \mathbb{P}(S_1 \geq \log(t)) = 0$. Conditional on $\Phi = 1$, $(H_j)_{j\geq 2}$ are independent therefore by Markov's inequality we have that

$$\mathbb{P} \left(\sum_{j=2}^{b_n^K \log(t)} \beta^{H_j - \overline{H}_n^K} S_j \geq t - \log(t) \,\middle|\, \bigcap_{j\geq 2} H_j \leq \overline{H}_n^K \right)$$

$$\leq \frac{b_n^K \log(t) \mathbb{E}[S_1] \mathbb{E}\left[\beta^{H_1} | H_1 \leq \overline{H}_n^K \right]}{\beta^{\overline{H}_n^K} (t - \log(t))}.$$

For large enough n we have that $\mathbb{P}(H_1 \leq \overline{H}_n^K) \geq 1/2$ therefore we have that

$$\mathbb{P}\left(H_1 = l \middle| H_1 \leq \overline{H}_n^K\right) \leq \mathbb{P}\left(H_1 \geq l \middle| H_1 \leq \overline{H}_n^K\right) \leq \frac{1}{2}\mathbb{P}\left(H_1 \geq l\right) \leq C\mu^l$$

for some constant C therefore the result follows from

$$\mathbb{E}\left[\beta^{H_1} \middle| H_1 \leq \overline{H}_n^K\right] = \sum_{l=0}^{\overline{H}_n^K} \beta^l \mathbb{P}\left(H_1 = l \middle| H_1 \leq \overline{H}_n^K\right) \leq C(\beta\mu)^{\overline{H}_n^K}. \qquad (7.27)$$

$\square$

We now prove three technical lemmas which will be important in the proof of Proposition 7.14 which is the main result of the section. The first shows that we can reintroduce the small traps into Z_∞^n. The reason for doing this is that we no longer need to condition on the heights of the traps being at least the critical level which will simplify later calculations. In particular, we can replace N with $\xi^* - 1$ (i.e. the total number of traps in the branch) which we understand under $\mathbf{P}^K$ by Lemma 7.5.

Lemma 7.11 *For all $\tilde{\varepsilon} > 0$ we have that for any $K \in \mathbb{Z}$ as $n \to \infty$,*

$$\mathbb{P}^K\left(\sum_{j=1}^{\xi^*-1-N} \beta^{H_j^- - \overline{H}_n^K} S_j > \tilde{\varepsilon}\right) \to 0.$$

Proof First, notice that each term in the sum is nonnegative therefore introducing extra terms only increases the probability. By Lemma 7.5, for any $\tilde{\epsilon} > 0$, we have that $\mathbb{P}^K(\xi^* - 1 \geq a_{n^{1+\tilde{\epsilon}}}) \to 0$ as $n \to \infty$. We therefore have that

$$\mathbb{P}^K\left(\sum_{j=1}^{\xi^*-1-N} \beta^{H_j^- - \overline{H}_n^K} S_j > \tilde{\varepsilon}\right) \leq \mathbb{P}\left(\sum_{j=1}^{a_{n^{1+\tilde{\epsilon}}}} \beta^{H_j - \overline{H}_n^K} S_j > \tilde{\varepsilon} \middle| H_j < h_{n,\varepsilon} \; \forall j \geq 1\right)$$
$$+ o(1).$$

By Definitions 7 and 8 we have that $\beta^{\overline{H}_n^K} \leq \beta^K a_n^{1/\gamma}$ therefore by Markov's inequality and (7.27) we have that

$$\mathbb{P}\left(\sum_{j=1}^{a_{n^{1+\tilde{\epsilon}}}} \beta^{H_j - \overline{H}_n^K} S_j > \tilde{\varepsilon} \middle| H_j < h_{n,\varepsilon} \; \forall j \geq 1\right) \leq \frac{a_{n^{1+\tilde{\epsilon}}} \mathbb{E}\left[S_1\right] \mathbb{E}\left[\beta^{H_1} \middle| H_1 < h_{n,\varepsilon}\right]}{\tilde{\varepsilon}\beta^{\overline{H}_n^K}}$$
$$\leq \frac{C_{K,\tilde{\varepsilon}} a_{n^{1+\tilde{\epsilon}}}(\beta\mu)^{h_{n,\varepsilon}}}{a_n^{1/\gamma}}.$$

Recall from Definition 7 that $h_{n,\varepsilon} \leq \log(a_{n^{1-\varepsilon}})/\log(\mu^{-1})$ therefore

$$(\beta\mu)^{h_{n,\varepsilon}} \leq a_{n^{1-\varepsilon}}^{\frac{1}{\gamma}-1}.$$

Using the form of a_n following Definition 3 we then have that there exists a slowly varying function $\overline{L}$ such that

$$\frac{a_{n^{1+\tilde{\varepsilon}}}(\beta\mu)^{h_{n,\varepsilon}}}{a_n^{1/\gamma}} \leq \overline{L}(n)n^{\frac{1}{\alpha-1}\left(\tilde{\varepsilon}+\varepsilon-\frac{\varepsilon}{\gamma}\right)}$$

which converges to 0 by choosing $\tilde{\varepsilon} < \varepsilon(1/\gamma - 1)$. $\qquad\square$

The second Lemma leading to Proposition 7.14 shows that the height of an f-GW-tree is sufficiently close to a geometric random variable. To ease notation let $S = S_1$ (see (7.24)), $H = H_1 \sim \mathcal{H}(\mathcal{T}^\circ)$ be distributed as the height of a GW-tree and $G \sim Geo(\mu)$ independently of each other.

Lemma 7.12 *In IVIE,*

$$b\left|\int_0^\infty e^{-x}\mathbb{P}\left(S\beta^H \geq \frac{xb^{1/\gamma}}{\theta}\right)dx - c_\mu \int_0^\infty e^{-x}\mathbb{P}\left(S\beta^G \geq \frac{xb^{1/\gamma}}{\theta}\right)dx\right| \quad (7.28)$$

converges to zero as $b \to \infty$.

Proof From (1.1) and (7.25) we have that $\gamma < 1$ and $\mathbb{E}[S] < \infty$ therefore $\mathbb{E}[S^\gamma] < \infty$. By independence of S and G

$$\mathbb{P}\left(S\beta^G \geq \frac{xb^{1/\gamma}}{\theta}\right) = \mathbb{E}\left[\mathbb{P}\left(G \geq \frac{\log\left(xb^{1/\gamma}(S\theta)^{-1}\right)}{\log(\beta)}\bigg| S\right)\right]$$

$$\leq \left(\frac{xb^{1/\gamma}}{\theta}\right)^{-\gamma}\mathbb{E}[S^\gamma] = \frac{C_\theta}{bx^\gamma}.$$

Similarly, since there exist a constant c such that $\mathbb{P}(H \geq t) \leq c\mathbb{P}(G \geq t)$ uniformly over t we have that

$$\mathbb{P}\left(S\beta^H \geq \frac{xb^{1/\gamma}}{\theta}\right) \leq \frac{C_\theta}{bx^\gamma}.$$

Let $\tilde{\varepsilon} > 0$ then choose $\epsilon > 0$ such that

$$\int_0^\epsilon e^{-x}x^{-\gamma}dx < \frac{\tilde{\varepsilon}}{C_\theta}$$

then, since the integrals are positive and $c_\mu \leq 1$, we have that

$$\left| \int_0^\epsilon e^{-x} c_\mu \mathbb{P}\left(S\beta^G \geq \frac{xb^{1/\gamma}}{\theta} \right) dx - \int_0^\epsilon e^{-x} \mathbb{P}\left(S\beta^H \geq \frac{xb^{1/\gamma}}{\theta} \right) dx \right| \leq \tilde{\epsilon} b^{-1}.$$

(7.29)

By (3.1) we have that

$$m(b) := \sup_{z > \frac{\epsilon b}{\theta}} \left| \frac{\mathbb{P}\left(\beta^H \geq z \right)}{\mathbb{P}\left(\beta^G \geq z \right)} - c_\mu \right|$$

(7.30)

converges to 0 as $b \to \infty$. Now define $M(b) := m(b)^{1-\frac{1}{\gamma}} \wedge b^{\frac{1}{\gamma}-1}$ then $M(b) \to \infty$ as $b \to \infty$ but $M(b) << b^{1/\gamma}$.

For $x > \epsilon$, by independence of S and H we have that

$$\mathbb{P}\left(S\beta^H \geq \frac{xb^{1/\gamma}}{\theta} \middle| S \geq M(b) \right) \leq C\mathbb{E}\left[\left(\frac{xb^{1/\gamma}}{\theta S} \right)^{\frac{\log(\mu)}{\log(\beta)}} \middle| S \geq M(b) \right]$$

$$\leq C_{\epsilon,\theta} b^{-1} \mathbb{E}\left[S^\gamma \middle| S \geq M(b) \right].$$

In particular,

$$b\mathbb{P}\left(S\beta^H \geq \frac{xb^{1/\gamma}}{\theta}, \ S \geq M(b) \right) \leq C_{\epsilon,\theta} \mathbb{E}\left[S^\gamma \mathbf{1}_{\{S \geq M(b)\}} \right]$$

which converges to 0 as $b \to \infty$ by dominated convergence. Similarly, the same holds replacing H with G therefore combining this with (7.29) we have that the quantity (7.28) is bounded above by

$$\tilde{\epsilon} + o(1) + Cb \sup_{x > \epsilon} \left| \mathbb{P}\left(S\beta^H \geq \frac{xb^{1/\gamma}}{\theta}, \ S < M(b) \right) - c_\mu \mathbb{P}\left(S\beta^G \geq \frac{xb^{1/\gamma}}{\theta}, \ S < M(b) \right) \right|.$$

Since S is independent of G and H we have that the supremum in the above expression can be bounded above by

$$\sup_{z > \frac{\epsilon b^{1/\gamma}}{\theta M(b)}} \left| \mathbb{P}\left(\beta^H \geq z \right) - c_\mu \mathbb{P}\left(\beta^G \geq z \right) \right| \leq m(b) \mathbb{P}\left(G \geq \frac{\log\left(\epsilon b^{1/\gamma} (\theta M(b))^{-1} \right)}{\log(\beta)} \right)$$

by (7.30) since $b^{1/\gamma}/M(b) \geq b$. Since $G \sim Geo(\mu)$ we have that

$$m(b)\mathbb{P}\left(G \geq \frac{\log(\epsilon b^{1/\gamma} (\theta M(b))^{-1})}{\log(\beta)} \right) = C_{\epsilon,\theta} m(b) \left(\frac{b^{1/\gamma}}{M(b)} \right)^{\frac{\log(\mu)}{\log(\beta)}} \leq \frac{C_{\epsilon,\theta} m(b)^\gamma}{b}$$

which completes the proof. $\qquad\square$

In the final Lemma preceding Proposition 7.14 we show that the Laplace transform

$$\varphi_K(\lambda) := \mathbb{E}^K\left[e^{-\lambda \sum_{j=1}^{\xi^*-1} \beta^{H_j - \overline{H}_n^K} S_j}\right]$$

can be written in terms of the distributions of S, H and ξ^*.

Lemma 7.13 *In IVIE,*

$$\varphi_K(\lambda) = \mathbb{E}^K\left[\frac{\mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}}1_{\{H \leq \overline{H}_n^K\}}\right]^{\xi^*-1} - \mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}}1_{\{H \leq \overline{H}_n^K - 1\}}\right]^{\xi^*-1}}{\mathbf{P}\left(H \leq \overline{H}_n^K\right)^{\xi^*-1} - \mathbf{P}\left(H \leq \overline{H}_n^K - 1\right)^{\xi^*-1}}\right]$$

Proof Recall that $\Phi := \inf\{r \geq 1 : H_r = \overline{H}\}$ is the index of the first random variable in the sequence $(H_j)_{j=1}^{\xi^*-1}$ which attains the maximum value $\overline{H} := \max_{j \leq \xi^*-1} H_j$. For $h \in \mathbb{Z}^+, \lambda > 0$ and $i = 1, 2$ write

$$\psi_i(h, \lambda) := \mathbb{E}\left[e^{-\lambda S\beta^{H-h}}\big| H \leq h + 1 - i\right],$$

$$\phi_i(h, \lambda) := \mathbb{E}\left[e^{-\lambda S\beta^{H-h}}1_{\{H \leq h+1-i\}}\right] = \psi_i(h, \lambda)\mathbb{P}(H \leq h+1-i). \quad (7.31)$$

Conditional on Φ, the random variables $(H_j)_{j \geq 1}$ are independent with

$$\mathbb{P}^K\left(H_j = z | \Phi\right) = \begin{cases} 1_{\{z = \overline{H}_n^K\}}, & \text{if } j = \Phi, \\ \mathbb{P}\left(H = z \big| H \leq \overline{H}_n^K - 1\right), & \text{if } j < \Phi, \\ \mathbb{P}\left(H = z \big| H \leq \overline{H}_n^K\right), & \text{if } j > \Phi. \end{cases}$$

By conditioning on ξ^*, we then have that

$$\varphi_K(\lambda) = \mathbb{E}^K\left[\sum_{k=1}^{\xi^*-1} \mathbf{P}^K(\Phi = k|\xi^*)\mathbb{E}^K\left[e^{-\lambda \sum_{j=1}^{\xi^*-1} \beta^{H_j - \overline{H}_n^K} S_j}\Big| \Phi = k, \xi^*\right]\right]$$

$$= \mathbb{E}^K\left[\mathbb{E}\left[e^{-\lambda S}\right]\sum_{k=1}^{\xi^*-1} \mathbf{P}^K(\Phi = k|\xi^*)\psi_2\left(\overline{H}_n^K, \lambda\right)^{k-1} \psi_1\left(\overline{H}_n^K, \lambda\right)^{\xi^*-1-k}\right]$$

$$(7.32)$$

and by Bayes' rule we also have that

$$\mathbf{P}^K(\Phi = k|\xi^*) = \frac{\mathbf{P}\left(H = \overline{H}_n^K\right)}{\mathbf{P}\left(\overline{H} = \overline{H}_n^K|\xi^*\right)}\mathbf{P}\left(H \le \overline{H}_n^K - 1\right)^{k-1}\mathbf{P}\left(H \le \overline{H}_n^K\right)^{\xi^*-1-k}.$$

(7.33)

Combining (7.31), (7.32) and (7.33) we can then write $\varphi_K(\lambda)$ as

$$\mathbb{E}^K\left[\frac{\mathbb{E}\left[e^{-\lambda S}\right]\mathbf{P}\left(H = \overline{H}_n^K\right)}{\mathbf{P}\left(\overline{H} = \overline{H}_n^K|\xi^*\right)}\sum_{k=1}^{\xi^*-1}\phi_2\left(\overline{H}_n^K,\lambda\right)^{k-1}\phi_1\left(\overline{H}_n^K,\lambda\right)^{\xi^*-1-k}\right]. \quad (7.34)$$

For $0 < p < q < 1$ and $l \in \mathbb{Z}^+$,

$$\sum_{k=1}^{l}p^{k-1}q^{l-k} = q^{l-1}\sum_{k=0}^{l-1}\left(\frac{p}{q}\right)^k = q^{l-1}\left(\frac{1-\left(\frac{p}{q}\right)^l}{1-\frac{p}{q}}\right) = \frac{q^l-p^l}{q-p}.$$

Since $0 < \phi_2(\overline{H}_n^K,\lambda)^{k-1} < \phi_1(\overline{H}_n^K,\lambda)^{k-1} < 1$, by (7.34) it follows that $\varphi_K(\lambda)$ is equal to

$$\mathbb{E}^K\left[\frac{\mathbb{E}\left[e^{-\lambda S}\right]\mathbf{P}\left(H = \overline{H}_n^K\right)}{\mathbf{P}\left(\overline{H} = \overline{H}_n^K|\xi^*\right)}\left(\frac{\phi_1\left(\overline{H}_n^K,\lambda\right)^{\xi^*-1} - \phi_2\left(\overline{H}_n^K,\lambda\right)^{\xi^*-1}}{\phi_1\left(\overline{H}_n^K,\lambda\right) - \phi_2\left(\overline{H}_n^K,\lambda\right)}\right)\right]$$

however, from (7.31),

$$\phi_1\left(\overline{H}_n^K,\lambda\right) - \phi_2\left(\overline{H}_n^K,\lambda\right) = \mathbb{E}\left[e^{-\lambda S}\right]\mathbf{P}\left(H = \overline{H}_n^K\right).$$

therefore this is equal to

$$\mathbb{E}^K\left[\frac{\phi_1\left(\overline{H}_n^K,\lambda\right)^{\xi^*-1} - \phi_2\left(\overline{H}_n^K,\lambda\right)^{\xi^*-1}}{\mathbf{P}\left(\overline{H} = \overline{H}_n^K|\xi^*\right)}\right]$$

The result then follows from

$$\mathbf{P}\left(\overline{H} = \overline{H}_n^K|\xi^*\right) = \mathbf{P}\left(\overline{H} \le \overline{H}_n^K|\xi^*\right) - \mathbf{P}\left(\overline{H} \le \overline{H}_n^K - 1|\xi^*\right)$$

$$= \mathbf{P}\left(H \le \overline{H}_n^K\right)^{\xi^*-1} - \mathbf{P}\left(H \le \overline{H}_n^K - 1\right)^{\xi^*-1}$$

which is a consequence of $\overline{H}$ being the maximum of $\xi^* - 1$ i.i.d. random variables. $\square$

The next proposition shows that, under $\mathbb{P}^K$, we have that the scaled time spent in a large branch $\zeta^{(n)}$ (from (7.18)) converges in distribution along subsequences n_l where $a_{n_l(t)} \sim t\mu^{-l}$.

Proposition 7.14 *In IVIE, under $\mathbb{P}^K$ we have that $Z_\infty^{n_l}$ converges in distribution (as $l \to \infty$) to some random variable Z_∞.*

Proof By Lemmas 7.11 and 7.13, it now suffices to show convergence of

$$\varphi_K(\lambda) = \mathbb{E}^K \left[\frac{\mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}} \mathbf{1}_{\left\{ H \leq \overline{H}_n^K \right\}} \right]^{\xi^*-1} - \mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}} \mathbf{1}_{\left\{ H \leq \overline{H}_n^K - 1 \right\}} \right]^{\xi^*-1}}{\mathbf{P}\left(H \leq \overline{H}_n^K \right)^{\xi^*-1} - \mathbf{P}\left(H \leq \overline{H}_n^K - 1 \right)^{\xi^*-1}} \right]. \tag{7.35}$$

By (3.2) we have that $\mathbf{P}(H \leq \overline{H}_n^K) = 1 - c_\mu \mu^{1+\overline{H}_n^K}(1+o(1))$ therefore, using the relationship (7.15) between b_n^K and $\overline{H}_n^K$ we have that

$$\mathbf{P}\left(H \leq \overline{H}_n^K \right)^{\xi^*-1} = \left(1 - \frac{\mu(1+o(1))}{b_n^K} \right)^{\xi^*-1} = \exp\left(-\frac{\xi^*-1}{b_n^K} \mu (1+o(1)) \right)$$

and similarly,

$$\mathbf{P}\left(H \leq \overline{H}_n^K - 1 \right)^{\xi^*-1} = \exp\left(-\frac{\xi^*-1}{b_n^K}(1+o(1)) \right).$$

By Lemma 7.5 we know that $(\xi^*-1)/b_n^K$ converges in distribution to a random variable with exponential moments therefore we want to show a similar expression for the numerator in (7.35). Notice that

$$\mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}} \mathbf{1}_{\left\{ H \leq \overline{H}_n^K \right\}} \right]^{\xi^*-1}$$

$$= \mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}} \right]^{\xi^*-1} \left(1 - \frac{\mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}} \mathbf{1}_{\left\{ H > \overline{H}_n^K \right\}} \right]}{\mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}} \right]} \right)^{\xi^*-1} \tag{7.36}$$

where $\mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}}\right]$ converges to 1 deterministically. In particular, this means that

$$\left(1 - \frac{\mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}}\mathbf{1}_{\left\{H>\overline{H}_n^K\right\}}\right]}{\mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}}\right]}\right)^{\xi^*-1}$$

$$= \exp\left(-(\xi^*-1)\mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}}\mathbf{1}_{\left\{H>\overline{H}_n^K\right\}}\right](1+o(1))\right). \qquad (7.37)$$

By summing over the possible values of H and using independence of S and H we have that

$$\mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}}\mathbf{1}_{\left\{H>\overline{H}_n^K\right\}}\right] = (1+o(1))\sum_{j=\overline{H}_n^K+1}^{\infty}(1-\mu)\mu^j\mathbb{E}\left[e^{-\lambda S\beta^{j-\overline{H}_n^K}}\right]$$

$$= (1+o(1))\mu^{\overline{H}_n^K+1}\sum_{j=0}^{\infty}(1-\mu)\mu^j\mathbb{E}\left[e^{-\lambda S\beta^{j+1}}\right].$$

Recalling $G \sim Geo(\mu)$ independently of S then writing $\varphi^{SG}(\lambda)$ to be the Laplace transform of $S\beta^G$ and using the relationship (7.15) between b_n^K and $\overline{H}_n^K$ we therefore have that (7.37) can be written as

$$\exp\left(-\frac{\xi^*-1}{b_n^K}\mu\varphi^{SH}(\lambda\beta)(1+o(1))\right). \qquad (7.38)$$

It remains to deal with $\mathbb{E}[e^{-\lambda S\beta^{H-\overline{H}_n^K}}]^{b_n^K}$. To ease notation, let us write $b := b_n^K = c_\mu^{-1}\mu^{-\overline{H}_n^K}$ and $\theta = \lambda c_\mu^{-1/\gamma}$ then

$$\mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}}\right]^{b_n^K} = \mathbb{E}\left[e^{-\theta S\beta^H b^{-1/\gamma}}\right]^b$$

$$= \left(\int_0^1 \mathbb{P}\left(e^{-\theta S\beta^H b^{-1/\gamma}} \geq y\right)dy\right)^b$$

$$= \left(1 - \int_0^1 \mathbb{P}\left(S\beta^H \geq -\frac{\log(y)b^{1/\gamma}}{\theta}\right)dy\right)^b$$

$$= \left(1 - \int_0^\infty e^{-x}\mathbb{P}\left(S\beta^H \geq \frac{xb^{1/\gamma}}{\theta}\right)dx\right)^b$$

$$= \left(1 - \int_0^\infty e^{-x}c_\mu\mathbb{P}\left(S\beta^G \geq \frac{xb^{1/\gamma}}{\theta}\right)dx\right)^b + o(1) \qquad (7.39)$$

where the final equality holds by Lemma 7.12. Since S and G are independent we have that

$$\mathbb{P}\left(S\beta^G \geq z\right) = \mathbb{E}\left[\mathbb{P}\left(G \geq \frac{\log(z/S)}{\log(\beta)} \Big| S\right)\right] = \mathbb{E}\left[\mu^{\left\lceil \frac{\log(z/S)}{\log(\beta)} \right\rceil}\right].$$

Writing

$$J(z) := \left\lceil \frac{\log(z)}{\log(\beta)} \right\rceil - \frac{\log(z)}{\log(\beta)} \quad \text{and} \quad I(z) := \mathbb{E}\left[S^\gamma \mu^{-\left\lfloor \frac{\log(S)}{\log(\beta)} + J(z) \right\rfloor + \frac{\log(S)}{\log(\beta)} + J(z)}\right]$$

we then have that

$$\mathbb{P}(S\beta^G \geq z) = \mu^{\frac{\log(z)}{\log(\beta)}} \mathbb{E}\left[S^\gamma \mu^{-\left\lfloor \frac{\log(S)}{\log(\beta)} + J(z) \right\rfloor + \frac{\log(S)}{\log(\beta)} + J(z)}\right] = z^{-\gamma} I(z)$$

where, from (7.25), we also have that $I(z) \leq \mathbb{E}[S^\gamma] < \infty$ since $\gamma < 1$ by (1.1). Moreover, $J(z) = J(zm^{\log(\beta)})$ and $I(z) = I(zm^{\log(\beta)})$ for all $z \in \mathbb{R}$, $m \in \mathbb{Z}$.

Substituting this back into (7.39) we have that

$$\mathbb{E}\left[e^{-\theta S\beta^G b^{-1/\gamma}}\right]^b = \left(1 - \theta^\gamma b^{-1} \int_0^\infty e^{-x} x^{-\gamma} I\left(\frac{xb^{1/\gamma}}{\theta}\right) dx\right)^b + o(1).$$

For $t > 0$, along sequences $n_l(t)$ such that $a_{n_l(t)} \sim t\mu^{-l}$ we have that $(b_n^K)^{1/\gamma} \sim Cl^{\log(\beta)}$ therefore, since I is bounded, we have that along subsequences $n_l(t)$

$$\int_0^\infty e^{-x} x^{-\gamma} I\left(\frac{xb^{1/\gamma}}{\theta}\right) dx$$

converges to some positive function of θ. In particular, we have that

$$\mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}}\right]^{b_n^K}$$

converges to some constant in the interval $(0, 1)$. Combining this with (7.36) and (7.38) we have that

$$\mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}} \mathbf{1}_{\left\{H \leq \overline{H}_n^K\right\}}\right]^{\xi^*-1} = \exp\left(-\frac{\xi^*-1}{b_n^K} \mu C_{\lambda,\beta}(1+o(1))\right)$$

for some constant $C_{\mu,\beta}$ depending on the distribution of S. Furthermore, the same arguments gives us that

$$\mathbb{E}\left[e^{-\lambda S\beta^{H-\overline{H}_n^K}} \mathbf{1}_{\left\{H \leq \overline{H}_n^K - 1\right\}}\right]^{\xi^*-1} = \exp\left(-\frac{\xi^*-1}{b_n^K} C_{\lambda,\beta}(1+o(1))\right).$$

By boundedness, continuity and Lemma 7.5 we therefore have that $\varphi_K(\lambda)$ converges along the given subsequences which proves the result. □

In order to prove the convergence result for sums of i.i.d. variables we shall require that $\zeta^{(n)}$ can be dominated (independently of $K \geq h_{n,\varepsilon} - h_{n,0}$) by some random variable Z_{sup} such that $\mathbb{E}[Z_{sup}^{(\alpha-1)\gamma+\epsilon}] < \infty$ for ϵ sufficiently small. Lemma 7.15 shows that we indeed have the domination required for the sums of i.i.d. variables result.

Lemma 7.15 *In IVIE, there exists a random variable Z_{sup} such that under $\mathbb{P}^K$ for any $K \in \mathbb{Z}$ we have that $Z_{sup} \succeq \zeta^{(n)}$ for all n sufficiently large and $\mathbb{E}[Z_{sup}^{1-\epsilon}] < \infty$ for any $\epsilon > 0$.*

Proof The number of large traps N is dominated by the number of traps in the branch. Similarly to Lemma 7.5 we consider

$$\mathbf{P}\left(\xi^* - 1 \geq tb_n^K \middle| \overline{H} = \overline{H}_n^K\right) = \mathbf{P}\left(\overline{H} = \overline{H}_n^K \middle| \xi^* - 1 \geq tb_n^K\right) \frac{\mathbf{P}\left(\xi^* - 1 \geq tb_n^K\right)}{\mathbf{P}\left(\overline{H} = \overline{H}_n^K\right)}.$$

Using the tail of H from (3.1), for large n (independently of $t \geq 0$) and some constant c, we can bound $\mathbf{P}(\overline{H} = \overline{H}_n^K | \xi^* - 1 \geq tb_n^K)$ above by

$$\mathbf{E}\left[\mathbf{P}\left(H \leq \overline{H}_n^K\right)^{\xi^*-1} \middle| \xi^* - 1 \geq tb_n^K\right] \leq \mathbf{E}\left[e^{-c\left(\frac{\xi^*-1}{b_n^K}\right)} \middle| \xi^* - 1 \geq tb_n^K\right] \leq e^{-ct}.$$

For each $t \geq 0$ we have that $\mathbf{P}(\xi^* - 1 \geq tb_n^K) \sim Ct^{-(\alpha-1)}\mathbf{P}(\overline{H} = \overline{H}_n^K)$ as $n \to \infty$. Since $\mathbf{P}(\overline{H} = \overline{H}_n^K)$ doesn't depend on t we can choose a constant c such that for n sufficiently large we have that $\mathbf{P}(\overline{H} = \overline{H}_n^K) \leq c\mathbf{P}(\xi^* - 1 \geq b_n^K)$ thus for $t \geq 1$

$$\frac{\mathbf{P}\left(\xi^* - 1 \geq tb_n^K\right)}{\mathbf{P}\left(\overline{H} = \overline{H}_n^K\right)} \leq \frac{\mathbf{P}\left(\xi^* - 1 \geq tb_n^K\right)}{c\mathbf{P}\left(\xi^* - 1 \geq b_n^K\right)} \leq c^{-1}.$$

In particular, for $t \geq 1$ we have that $\mathbf{P}(\xi^* - 1 \geq tb_n^K | \overline{H} = \overline{H}_n^K) \leq c_1 e^{-c_2 t}$ for some constants c_1, c_2. It follows that there exists some random variable ξ_{sup} which is independent of $\overline{H}$, has an exponential tail and satisfies $\xi_{sup}b_n^K \geq \xi^* - 1$ on the event $\{\overline{H} = \overline{H}_n^K\}$ for n suitably large (independently of K).

Recall that the total number of excursions W^j in a trap exceeds the number which reach the deepest point B^j and we write $G^{j,k}$ to denote the number of excursions from the deepest point. The length of these excursions can be dominated by excursions $\mathcal{R}_\infty^{j,k,l}$ from the deepest points of the infinite traps $\mathcal{T}_i^\prec$. We then have that for n suitably large, under $\mathbb{P}^K$

$$\zeta^{(n)} \preceq \sum_{j=1}^{\xi_{sup}b_n^K} \sum_{k=1}^{W^j} \sum_{l=1}^{G^{j,k}} \frac{\mathcal{R}_\infty^{j,k,l}}{\beta^{\overline{H}_n^K}}.$$

By (7.20) $E[G^{j,k}] \leq \beta^{H_j+1}/(\beta + 1)$ therefore there is some constant c such that, writing

$$\mathcal{Y}_j^{(n)} := c \sum_{k=1}^{W^j} \sum_{l=1}^{G^{j,k}} \frac{\mathcal{R}_\infty^{j,k,l}}{E[G^{j,k}]}$$

(which are identically distributed under $\mathbb{P}$) we have that under $\mathbb{P}^K$,

$$\zeta^{(n)} \preceq \frac{1}{\beta^{\overline{H}_n^K}} \sum_{j=1}^{\xi_{sup}b_n^K} \beta^{H_j} \mathcal{Y}_j^{(n)}.$$

For $m \geq 1$ write $\mathscr{X}^n(m) := \frac{1}{m} \sum_{j=1}^m \beta^{H_j} \mathcal{Y}_j^{(n)} \mathbf{1}_{\{j \neq \Phi\}}$ (where we recall that Φ is the first index j such that $H_j = \overline{H}_n^K$) then by Markov's inequality

$$\mathbb{P}^K \left(\mathscr{X}^n(m) \geq t \right) \leq \frac{1}{m} \sum_{j=1}^m \frac{\mathbb{E}^K \left[\beta^{H_j} \mathcal{Y}_j^{(n)} \mathbf{1}_{\{j \neq \Phi\}} \right]}{t}$$

$$= \frac{1}{m} \sum_{j=1}^m \frac{\mathbb{E}^K \left[\beta^{H_j} \mathbf{1}_{\{j \neq \Phi\}} \right] \mathbb{E}^K \left[\mathcal{Y}_j^{(n)} \right]}{t}$$

since $\mathbb{E}[\mathcal{Y}_j^{(n)} | H_j, \Phi]$ is independent of H_j and Φ. Since W^1 has a geometric distribution (independently of n) we have that $\mathbb{E}[W^1] < \infty$ and by Lemma 7.1 we have that $\mathbb{E}[\mathcal{R}_\infty] < \infty$ therefore $\mathbb{E}^K[\mathcal{Y}_j^{(n)}] \leq \mathbb{E}[W^1]\mathbb{E}[\mathcal{R}_\infty] \leq C < \infty$ for all n. Using geometric bounds on the tail of H from (3.1) and that $\mathbb{P}(H \geq j | H \leq \overline{H}_n^K) \leq \mathbb{P}(H \geq j)$ we have that

$$\mathbb{E}^K \left[\beta^H \right] \leq \sum_{j=0}^\infty \beta^j \mathbb{P} \left(H \geq j \big| H \leq \overline{H}_n^K \right) \leq C(\beta\mu)^{\overline{H}_n^K}.$$

We therefore have that $\mathbb{P}^K(\mathscr{X}^n(m) \geq t) \leq C(\beta\mu)^{\overline{H}_n^K}/t$ thus there exists some sequence of random variables $\mathscr{X}_{sup}^n \succeq \mathscr{X}^n(m)$ for any m such that $\mathbb{P}^K(\mathscr{X}_{sup}^n \geq t) = 1 \wedge C(\beta\mu)^{\overline{H}_n^K} t^{-1}$. In particular, $\mathscr{X}_{sup}^n \succeq \mathscr{X}^n(\xi_{sup}b_n^K)$. Therefore,

$$\frac{1}{\beta^{\overline{H}_n^K}} \sum_{j=1}^{\xi_{sup}b_n^K} \beta^{H_j} \mathcal{Y}_j^{(n)} = \frac{\xi_{sup}b_n^K}{\beta^{\overline{H}_n^K}} \mathscr{X}^n(\xi_{sup}b_n^K) + \mathcal{Y}_\Phi^{(n)} \preceq \frac{\xi_{sup} \mathscr{X}_{sup}^n}{c_\mu(\beta\mu)^{\overline{H}_n^K}} + \mathcal{Y}_\Phi^{(n)}$$

under $\mathbb{P}^K$. We then have that

$$\mathbb{P}^K\left(\frac{\xi_{sup}\mathscr{X}^n_{sup}}{c_\mu(\beta\mu)^{\overline{H}^K_n}} \geq t\right) = \mathbb{E}^K\left[\mathbb{P}^K\left(\mathscr{X}^n_{sup} \geq \frac{tc_\mu(\beta\mu)^{\overline{H}^K_n}}{\xi_{sup}}\bigg|\xi_{sup}\right)\right] = 1 \wedge C\frac{\mathbb{E}^K\left[\xi_{sup}\right]}{t}$$

where ξ_{sup} has finite first moment since $\mathbf{P}(\xi_{sup} \geq t) = c_1e^{-c_2t} \wedge 1$. It follows that there exists $\mathscr{X}_{sup} \succeq \mathscr{X}^n_{sup}$ for any n such that $\mathbb{P}(\mathscr{X}_{sup} \geq t) = 1 \wedge Ct^{-1}$.

Since $\mathbb{E}^K[\mathscr{Y}^n_\Phi]$ is bounded independently of K and n, by Markov's inequality we have that there exists $\mathscr{Y}_{sup} \succeq \mathscr{Y}^n_\Phi$ for all n such that $\mathbb{P}(\mathscr{Y}_{sup} \geq t) = 1 \wedge Ct^{-1}$. It therefore follows that $\zeta^{(n)}$ under $\mathbb{P}^K$ is stochastically dominated by $\mathscr{X}_{sup} + \mathscr{Y}_{sup}$ under $\mathbb{P}$ where

$$\mathbb{P}\left(\mathscr{X}_{sup} + \mathscr{Y}_{sup} \geq t\right) \leq \mathbb{P}\left(\mathscr{X}_{sup} \geq t/2\right) + \mathbb{P}\left(\mathscr{Y}_{sup} \geq t/2\right) \leq Ct^{-1}$$

hence $\mathscr{X}_{sup} + \mathscr{Y}_{sup}$ has finite moments up to $1 - \epsilon$ for all $\epsilon > 0$. □

8 Convergence along subsequences

In this section we prove the main theorems concerning convergence to infinitely divisible laws in FVIE and IVIE. Both cases follow the proof from [4]; in FVIE the result follows directly whereas in IVIE adjustments need to be made to deal with slowly varying functions.

Recall that we want to show convergence of Δ_n/a_n along sequences $n_l(t)$ however by Corollary 5.5 and Lemma 5.6 it suffices to consider

$$\tilde{\chi}_{t,n} = \sum_{i=1}^{\lfloor ntq_n \rfloor} \tilde{\chi}^i_n$$

where $\tilde{\chi}^i_n$ is the time spent in large traps of the ith large branch by walk $X^{(i)}_n$. Furthermore, by Proposition 7.3 and Corollary 7.4 we can replace $\tilde{\chi}^i_n$ with $\tilde{\chi}^{i*}_n$ which is the time spent on excursions from the deepest point of the traps of the ith branch by $X^{(i)}_n$.

Let $\overline{H}_i$ denote the height of the largest trap in the ith large branch then for $i, l \geq 1$ let $\zeta^l_i := \tilde{\chi}^{i*}_{n_l}\beta^{-\overline{H}_i}$ then $(\zeta^l_i)_{i\geq 1}$ are i.i.d. with the law of $\zeta^{(n_l)}$. Let $n_l := n_l(1)$ then for $K \geq -(l - h_{n_l,\varepsilon})$ let $\zeta^{l,K}_i$ be ζ^l_i conditioned on the event $\{\overline{H}_i = l + K\}$ when this makes sense and 0 otherwise. For $K \in \mathbb{Z}$ and $l \geq 0$ define $\overline{F}^l_K(x) := \mathbb{P}(\zeta^{l,K}_i > x)$.

8.1 Proof of Theorem 2 (FVIE)

Recall that in FVIE $\gamma = \log(\mu^{-1})/\log(\beta) < 1$, $n_l(t) = \lfloor t\mu^{-l}\rfloor$ and by Corollary 3.2 we have that the height of a branch decays exponentially: $\mathbf{P}(\mathcal{H}(\mathcal{T}^{*-}) \geq n) \sim C_\mathcal{D}\mu^n = C_\mathcal{D}\beta^{-n\gamma}$ where $C_\mathcal{D} = c_\mu\mathbf{E}[\xi^* - 1]$.

By a simple adaptation of Corollary 7.10 and Lemma 7.15

1. $\exists Z_\infty^{(i)}$ random variables such that for all $K \in \mathbb{Z}$ we have that $\zeta_i^{l,K} \xrightarrow{d} Z_\infty^{(i)}$ as $l \to \infty$;
2. $\exists Z_{sup}$ random variable such that for all $l \geq 0$ and $K \geq -(l - h_{n_l,\varepsilon})$ we have that $\zeta_i^{l,K} \preceq Z_{sup}$ and $\mathbf{E}[Z_{sup}^{\gamma+\epsilon}] < \infty$ for some $\epsilon > 0$.

More specifically, since there is precisely one large trap in a large branch with high probability in FVIE the random variable in (7.23) can be written as

$$Z_\infty^n = \frac{1}{1 - \beta^{-1}} E[\mathcal{R}_\infty] \sum_{k=1}^{B_\infty} e_k$$

for some binomial variable B_∞ and independent exponential variables e_k. These are independent of n, hence an adaptation of Proposition 7.10 shows that $\zeta^{(n)}$ converge in distribution under $\mathbb{P}^K$.

Set

$$S_M^l := \sum_{i=1}^M \tilde{\chi}_{n_l}^{i*}.$$

For $(\lambda_l)_{l \geq 0}$ converging to $\lambda > 0$ define $M_l^\lambda := \lfloor \lambda_l^\gamma \beta^{\gamma(l - h_{n_l,\varepsilon})} \rfloor$ and $K_l^\lambda := \lambda \beta^l$ then denote $\overline{F}_\infty(x) := \mathbf{P}(Z_\infty > x)$. Theorem 7 is Theorem 10.1 of [4].

Theorem 7 *Suppose $\gamma < 1$ and properties 1 and 2 hold then*

$$S_{M_l^\lambda}^l / K_l^\lambda \xrightarrow{d} R_{d_\lambda, 0, \mathcal{L}_\lambda}$$

where $R_{d_\lambda, 0, \mathcal{L}_\lambda}$ has an infinitely divisible law with drift

$$d_\lambda = \lambda^{1+\gamma} \left(1 - \beta^{-\gamma}\right) \sum_{K \in \mathbb{Z}} \beta^{(1+\gamma)K} \mathbf{E}\left[\frac{Z_\infty}{(\lambda \beta^K)^2 + (Z_\infty)^2}\right],$$

0 variance and Lévy spectral function $\mathcal{L}_\lambda$ satisfying $\mathcal{L}_\lambda(x) = \lambda^\gamma \mathcal{L}_1(\lambda x)$ for all $\lambda > 0, x \in \mathbb{R}$ with $\mathcal{L}_1(x) = 0$ for $x < 0$ and

$$\mathcal{L}_1(x) = -\left(1 - \beta^{-\gamma}\right) \sum_{K \in \mathbb{Z}} \beta^{K\gamma} \overline{F}_\infty(x\beta^K)$$

for $x \geq 0$.

Combining this with the remark at the beginning of the section with $\lambda = (tC_\mathcal{D})^{1/\gamma} = (tc_\mu \mathbf{E}[\xi^* - 1])^{1/\gamma}$ and that eventually $l = h_{n_l,0}$ we have that

$$\frac{\Delta_{n_l}(t)}{(C_\mathcal{D} n_l(t))^{\frac{1}{\gamma}}} \xrightarrow{d} R_{d_{(tC_\mathcal{D})^{1/\gamma}}, 0, \mathcal{L}_{(tC_\mathcal{D})^{1/\gamma}}}$$

which proves Theorem 2.

8.2 Proof of Theorem 3 (IVIE)

In IVIE write $\gamma_\alpha := (\alpha - 1)\log(\mu^{-1})/\log(\beta) = (\alpha - 1)\gamma$. By (3.12) we have that

$$\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) > n\right) \sim c_\mu^{\alpha-1}\Gamma(2 - \alpha)\mathbf{P}\left(\xi^* \geq \mu^{-n}\right) \sim C_{\mu,\alpha}\beta^{-\gamma_\alpha n}L\left(\beta^{\gamma_\alpha n}\right)$$

for a known constant $C_{\mu,\alpha}$. Due to the slowly varying term, we cannot apply Theorem 7 directly however Theorem 7 is proved using Theorem 6. It will therefore suffice to show convergence of the drift, variance and Lévy spectral function in this case.

Recall that we consider subsequences $n_l(t)$ such that $a_{n_l(t)} \sim t\mu^{-l}$. From Propositions 7.10 and 7.14 we then have that for any $K \in \mathbb{Z}$ the laws of $\zeta_i^{l,K}$ converge to the laws of Z_∞ as $l \to \infty$. Let $(Z_\infty^{(i)})_{i\geq 1}$ be an independent sequence of variables with this law and denote $\overline{F}_\infty(x) := \mathbb{P}(Z_\infty > x)$. By Lemma 7.15, $\exists Z_{sup}$ such that $\zeta_i^{l,K} \preceq Z_{sup}$ for all $l \in \mathbb{N}$, $K \geq -(l - h_{n_l,\varepsilon})$ and $\mathbf{E}[Z_{sup}^{\gamma_\alpha+\epsilon}] < \infty$ for some $\epsilon > 0$; we denote $\overline{F}_{sup}(x) := \mathbf{P}(Z_{sup} > x)$. For $(\lambda_l)_{l\geq 0}$ converging to $\lambda > 0$ define $K_l^\lambda := \lambda\beta^l$ and for $C_{\alpha,\mu} = \mu^{-1}(2 - \alpha)/(\alpha - 1)$

$$M_l^\lambda := \left\lfloor \lambda_l^{\gamma_\alpha}\beta^{\gamma_\alpha l}\frac{\mathbf{P}\left(\xi^* > \mu^{-h_{n_l,\varepsilon}}\right)}{C_{\alpha,\mu}L\left(\mu^{-h_{n_l,0}}\right)} \right\rfloor.$$

Proposition 8.1 *In IVIE, for any $\lambda > 0$, as $l \to \infty$*

$$\sum_{i=1}^{M_l^\lambda} \frac{\tilde{\chi}_{n_l}^{i*}}{K_l^\lambda} \xrightarrow{d} R_{d_\lambda,0,\mathcal{L}_\lambda}$$

where

$$d_\lambda = \lambda^{1+\gamma_\alpha}(1 - \beta^{-\gamma_\alpha})\sum_{K\in\mathbb{Z}}\beta^{(1+\gamma_\alpha)K}\mathbb{E}\left[\frac{Z_\infty}{(\lambda\beta^K)^2 + (Z_\infty)^2}\right],$$

$$\mathcal{L}_\lambda(x) = \begin{cases} 0 & x \leq 0; \\ -\lambda^{\gamma_\alpha}(1 - \beta^{-\gamma_\alpha})\sum_{K\in\mathbb{Z}}\beta^{K\gamma_\alpha}\overline{F}_\infty(\lambda x\beta^{K\gamma_\alpha}) & x > 0. \end{cases}$$

Proof By Theorem 6 it suffices to show the following:

1. for all $\epsilon > 0$

$$\lim_{l\to\infty} \mathbb{P}\left(\frac{\tilde{\chi}_{n_l}^{1*}}{K_l^\lambda} > \epsilon\right) = 0;$$

2. for all x continuity points

$$
\mathcal{L}_\lambda(x) = \begin{cases} 0 & x \leq 0, \\ -\lim_{l\to\infty} M_l^\lambda \mathbb{P}\left(\frac{\tilde{\chi}_{n_l}^{1*}}{K_l^\lambda} > x\right) & x > 0; \end{cases}
$$

3. for all $\tau > 0$ continuity points of $\mathcal{L}$

$$
d_\lambda = \lim_{l\to\infty} M_l^\lambda \mathbb{E}\left[\frac{\tilde{\chi}_{n_l}^{1*}}{K_l^\lambda} \mathbf{1}_{\left\{\tilde{\chi}_{n_l}^{1*} \leq \tau K_l^\lambda\right\}}\right] + \int_{|x|\geq\tau} \frac{x}{1+x^2} d\mathcal{L}_\lambda(x)
$$
$$
- \int_{\tau\geq|x|>0} \frac{x^3}{1+x^2} d\mathcal{L}_\lambda(x);
$$

4.

$$
\lim_{\tau\to 0} \limsup_{l\to\infty} M_l^\lambda Var\left(\frac{\tilde{\chi}_{n_l}^{1*}}{K_l^\lambda} \mathbf{1}_{\left\{\tilde{\chi}_{n_l}^{1*} \leq \tau K_l^\lambda\right\}}\right) = 0.
$$

We prove each of these in turn but we start by introducing a relation which will be fundamental to proving the final parts. For $K \in \mathbb{Z}$ let $c_l^K = \mathbf{P}(\mathcal{H}(\mathcal{T}^{*-}) > l + K | \mathcal{H}(\mathcal{T}^{*-}) > h_{n_l,\varepsilon})$ denote the probability that a deep branch is of height at least $l + K$. Then by the asymptotic (3.12) we have that, for K such that $l + K \geq h_{n_l,\varepsilon}$, as $l \to \infty$

$$
c_l^K = \frac{\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) > l + K\right)}{\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) > h_{n_l,\varepsilon}\right)} \sim \mu^{(\alpha-1)K} \frac{\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) > l\right)}{\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) > h_{n_l,\varepsilon}\right)}.
$$

In particular, using (3.12) and that $\beta^{\gamma_\alpha} = \mu^{-(\alpha-1)}$

$$
M_l^\lambda c_l^K \sim \lambda^{\gamma_\alpha} \left(\frac{\mathbf{P}\left(\xi^* > \mu^{h_{n_l,\varepsilon}}\right)}{\mathbf{P}\left(\mathcal{H}\left(\mathcal{T}^{*-}\right) > h_{n_l,\varepsilon}\right)}\right) \left(\frac{\mathbf{P}\left(\mathcal{H}\left(\mathcal{T}^{*-}\right) > h_{n_l,0}\right)}{\mathbf{P}\left(\xi^* > \mu^{h_{n_l,0}}\right)}\right)
$$
$$
\frac{\beta^{-\gamma_\alpha l}}{\mu^{-(\alpha-1)l}} \mu^{(\alpha-1)K} \sim \lambda^{\gamma_\alpha} \beta^{-\gamma_\alpha K}
$$

thus $M_l^\lambda(c_l^K - c_l^{K+1}) \to \lambda^{\gamma_\alpha} \beta^{-\gamma_\alpha K}(1 - \beta^{-\gamma_\alpha})$ and for any $\epsilon > 0$ and large enough l

$$
M_l^\lambda c_l^K \leq C_\epsilon \lambda^{\gamma_\alpha} \beta^{-\gamma_\alpha K} \beta^{\epsilon|K|}. \tag{8.1}
$$

To prove (1), notice that

$$
\mathbb{P}\left(\frac{\tilde{\chi}_{n_l}^{1*}}{K_l^\lambda} > \epsilon\right) \leq \mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) \geq h_{n,\varepsilon/2} \big| \mathcal{H}(\mathcal{T}^{*-}) \geq h_{n,\varepsilon}\right) + \mathbb{P}\left(\beta^{h_{n,\varepsilon/2}-l} Z_{sup} > \lambda\epsilon\right).
$$

Both terms converge to 0 as $l \to \infty$ by the tail formula of a branch (3.12), the fact that Z_{sup} has no atom at ∞ and that $\beta^{h_{n_l}^{\varepsilon/2} - l} \to 0$ which follows from $l \sim h_{n,0}$.

For (2), recall that $F_K^l(x) = \mathbb{P}(\zeta_i^{l,K} > x) = \mathbb{P}^K(\tilde{\chi}_{n_l}^{1*} \beta^{-(l+K)} > x)$, therefore

$$M_l^\lambda \mathbb{P}\left(\frac{\tilde{\chi}_{n_l}^{1*}}{K_l^\lambda} > x\right) = \sum_{K \in \mathbb{Z}} \mathbf{1}_{\{K \geq -(l - h_{n_l, \varepsilon})\}} M_l^\lambda \mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) = l + K\right) \mathbb{P}^K\left(\frac{\tilde{\chi}_{n_l}^{1*}}{K_l^\lambda} > x\right)$$

$$= \sum_{K \in \mathbb{Z}} \mathbf{1}_{\{K \geq -(l - h_{n_l, \varepsilon})\}} M_l^\lambda \left(c_l^K - c_l^{K+1}\right) \overline{F}_K^l(\lambda \beta^{-K} x).$$

If $x > 0$ is a continuity point of $\mathcal{L}_\lambda$ then $\lambda x \beta^{-K}$ is a continuity point of $\overline{F}_\infty$ hence for any $K \in \mathbb{Z}$ as $l \to \infty$

$$\mathbf{1}_{\{K \geq -(l - h_{n_l, \varepsilon})\}} M_l^\lambda \left(c_l^K - c_l^{K+1}\right) \overline{F}_K^l(\lambda \beta^{-K} x) \to \lambda^{\gamma_\alpha} \beta^{-\gamma_\alpha K} (1 - \beta^{-\gamma_\alpha}) \overline{F}_\infty(\lambda \beta^{-K} x).$$

We need to exchange the sum and the limit; we do this using dominated convergence. Since $\gamma_\alpha < 1$ we can choose $\epsilon > 0$ such that $\gamma_\alpha + \epsilon < 1$ and $\epsilon < \gamma_\alpha$. By (8.1), for l sufficiently large $M_l^\lambda c_l^K \leq C_{\epsilon, \lambda} \beta^{-\gamma_\alpha K} \beta^{\frac{\epsilon}{2}|K|}$ hence

$$\sum_{K \geq -(l - h_{n_l, \varepsilon})} M_l^\lambda \left(c_l^K - c_l^{K+1}\right) \overline{F}_K^l\left(\lambda \beta^{-K} x\right) \leq C \sum_{K \in \mathbb{Z}} \overline{F}_{sup}\left(\lambda x \beta^{-K}\right) \beta^{-\gamma_\alpha K} \beta^{\frac{\epsilon}{2}|K|}.$$

Since Z_{sup} has moments up to $\gamma_\alpha + \epsilon$ we have that for $y = \lambda x$

$$\sum_{K < 0} \overline{F}_{sup}(\lambda x \beta^{-K}) \beta^{-\gamma_\alpha K} \beta^{\frac{\epsilon}{2}|K|} = \mathbb{E}\left[\sum_{K=0}^{\left\lfloor \frac{\log(Z_{sup}/y)}{\log(\beta)} \right\rfloor} \beta^{K(\gamma_\alpha + \epsilon/2)}\right] \leq C_y \mathbb{E}\left[Z_{sup}^{\gamma_\alpha + \frac{\epsilon}{2}}\right]$$

which is finite. By choice of ϵ it follows that

$$\sum_{K \geq 0} \overline{F}_{sup}(\lambda x \beta^{-K}) \beta^{-\gamma_\alpha K} \beta^{\frac{\epsilon}{2}|K|} \leq \sum_{K \geq 0} \beta^{\left(\frac{\epsilon}{2} - \gamma_\alpha\right)K} < \infty.$$

It therefore follows that for $x > 0$

$$-\lim_{l \to \infty} M_l^\lambda \mathbb{P}\left(\frac{\tilde{\chi}_{n_l}^{1*}}{K_l^\lambda} > x\right) = -\lambda^{\gamma_\alpha} \left(1 - \beta^{-\gamma_\alpha}\right) \sum_{K \in \mathbb{Z}} \overline{F}_\infty\left(\lambda x \beta^{\gamma_\alpha K}\right) \beta^{\gamma_\alpha K}.$$

Moreover, for $x < 0$ we have that $\mathbb{P}\left(\tilde{\chi}_{n_l}^{1*}/K_l^\lambda < x\right) = 0$ which gives (2).

For (3) we have that $\int_0^\tau x \, d\mathcal{L}_\lambda$ is well defined therefore

$$\int_\tau^\infty \frac{x}{1 + x^2} d\mathcal{L}_\lambda - \int_0^\tau \frac{x^3}{1 + x^2} d\mathcal{L}_\lambda = \int_0^\infty \frac{x}{1 + x^2} d\mathcal{L}_\lambda - \int_0^\tau x \, d\mathcal{L}_\lambda.$$

We therefore want to show that

$$\lim_{l \to \infty} \frac{M_l^\lambda}{K_l^\lambda} \mathbb{E}\left[\tilde{\chi}_{n_l}^{1*} \mathbf{1}_{\left\{ \tilde{\chi}_{n_l}^{1*} \leq \tau K_l^\lambda \right\}} \right] = \int_0^\tau x \, \mathrm{d}\mathcal{L}_\lambda.$$

Write $G_K^l(u) = \mathbb{E}\left[\zeta_1^{l,K} \mathbf{1}_{\left\{ \zeta_1^{l,K} \leq u \right\}} \right]$ and $G_\infty(u) = \mathbb{E}[Z_\infty \mathbf{1}_{\{Z_\infty \leq u\}}]$. Then we have that

$$\frac{M_l^\lambda}{K_l^\lambda} \mathbb{E}\left[\tilde{\chi}_{n_l}^{1*} \mathbf{1}_{\left\{ \tilde{\chi}_{n_l}^{1*} \leq \tau K_l^\lambda \right\}} \right] = \lambda^{-1} \sum_{K \geq -(l - h_{n_l, \varepsilon})} M_l^\lambda \left(c_l^K - c_l^{K+1} \right) \beta^K G_K^l(\tau \lambda \beta^{-K}).$$

For each $K \in \mathbb{Z}$ as $l \to \infty$

$$M_l^\lambda \left(c_l^K - c_l^{K+1} \right) \beta^K G_K^l(\tau \lambda \beta^{-K}) \to \lambda^{\gamma_\alpha} (1 - \beta^{-\gamma_\alpha}) \beta^{(1-\gamma_\alpha)K} G_\infty(\tau \lambda \beta^{-K}).$$

We want to exchange the limit and the sum which we do by dominated convergence. For any $\kappa \in [0, 1]$ and random variable Y we have that $\mathbb{E}[Y \mathbf{1}_{\{Y \leq u\}}] \leq u^\kappa \mathbb{E}[Y^{1-\kappa} \mathbf{1}_{\{Y \leq u\}}]$. Using this with $u = \tau \lambda \beta^{-K}$ where $\kappa = 1 - \gamma_\alpha - 2\epsilon/3$ for $K < 0$ and $\kappa = 1$ for $K \geq 0$, alongside (8.1) we have that

$$\sum_{K \in \mathbb{Z}} \mathbf{1}_{\left\{ K \geq -(l - h_{n_l, \varepsilon}) \right\}} M_l^\lambda \left(c_l^K - c_l^{K+1} \right) \beta^K G_K^l(\tau \lambda \beta^{-K})$$

$$\leq \sum_{K \geq 0} M_l^\lambda \left(c_l^K - c_l^{K+1} \right) \beta^K \tau \lambda \beta^{-K}$$

$$+ \sum_{K < 0} M_l^\lambda \left(c_l^K - c_l^{K+1} \right) \beta^K \left(\beta^{\frac{2\epsilon}{3}K} (\tau \lambda)^{1 - \gamma_\alpha - \frac{2\epsilon}{3}} \mathbb{E}\left[Z_{sup}^{\gamma_\alpha + \frac{2\epsilon}{3}} \right] \beta^{(\gamma_\alpha - 1)K} \right)$$

$$\leq C_\lambda \tau \sum_{K \geq 0} \beta^{-(\gamma_\alpha - \epsilon/2)K} + C_\lambda \tau^{1 - \gamma_\alpha - \frac{2\epsilon}{3}} \mathbb{E}\left[Z_{sup}^{\gamma_\alpha + \frac{2\epsilon}{3}} \right] \sum_{K < 0} \beta^{\frac{\epsilon}{6}K}$$

which is finite since $\gamma_\alpha > \epsilon/2$ and Z_{sup} has moments up to $\gamma_\alpha + \epsilon$. We therefore have that

$$\lim_{l \to \infty} \frac{M_l^\lambda}{K_l^\lambda} \mathbb{E}\left[\tilde{\chi}_{n_l}^{1*} \mathbf{1}_{\left\{ \tilde{\chi}_{n_l}^{1*} \leq \tau K_l^\lambda \right\}} \right] = \lambda^{\gamma_\alpha - 1} (1 - \beta^{-\gamma_\alpha}) \sum_{K \in \mathbb{Z}} \beta^{K(\gamma_\alpha - 1)} G_\infty(\tau \lambda \beta^K).$$

By definition we have that

$$\int_0^\tau x d\mathcal{L}_\lambda = \lambda^{\gamma_\alpha}(1-\beta^{-\gamma_\alpha})\int_0^\tau x \sum_{K\in\mathbb{Z}} \beta^{\gamma_\alpha K} d(-\overline{F}_\infty)(\lambda x\beta^K)$$

$$= \lambda^{\gamma_\alpha-1}(1-\beta^{-\gamma_\alpha})\sum_{K\in\mathbb{Z}} \beta^{(\gamma_\alpha-1)K}\int_{\lambda x\beta^K\le\lambda\tau\beta^K} \lambda x\beta^K d(-\overline{F}_\infty)(\lambda x\beta^K)$$

$$= \lambda^{\gamma_\alpha-1}(1-\beta^{-\gamma_\alpha})\sum_{K\in\mathbb{Z}} \beta^{(\gamma_\alpha-1)K} G_\infty(\tau\lambda\beta^K).$$

It therefore remains to calculate $\int_0^\infty \frac{x}{1+x^2} d\mathcal{L}_\lambda$.

$$\int_0^\infty \frac{x}{1+x^2} d\mathcal{L}_\lambda = \lambda^{\gamma_\alpha}(1-\beta^{-\gamma_\alpha})\int_0^\infty \frac{x}{1+x^2}\sum_{K\in\mathbb{Z}} \beta^{\gamma_\alpha K} d(-\overline{F}_\infty)(\lambda x\beta^K)$$

$$= \lambda^{\gamma_\alpha+1}(1-\beta^{-\gamma_\alpha})\sum_{K\in\mathbb{Z}} \beta^{(\gamma_\alpha+1)K}\mathbb{E}\left[\frac{Z_\infty}{(\lambda\beta^K)^2+(Z_\infty)^2}\right].$$

The final sum is finite since for $K<0$

$$\beta^{(\gamma_\alpha+1)K}\mathbb{E}\left[\frac{Z_\infty}{(\lambda\beta^K)^2+(Z_\infty)^2}\right] = \lambda^{-1}\beta^{\gamma_\alpha K}\mathbb{E}\left[\frac{\lambda\beta^K Z_\infty}{(\lambda\beta^K)^2+(Z_\infty)^2}\right] \le \lambda^{-1}\beta^{\gamma_\alpha K}$$

Which is summable and for $K\ge 0$

$$\mathbb{E}\left[\frac{Z_\infty}{(\lambda\beta^K)^2+(Z_\infty)^2}\right] \le \mathbb{E}\left[\frac{Z_\infty}{(\lambda\beta^K)^2}\mathbf{1}_{\{Z_\infty\le\lambda\beta^K\}} + Z_\infty^{-1}\mathbf{1}_{\{Z_\infty\ge\lambda\beta^K\}}\right]$$

$$\le C_\lambda\mathbb{E}\left[Z_{sup}^{\gamma_\alpha+\epsilon/2}\right]\beta^{-K(1+\gamma_\alpha+\epsilon/2)}$$

which, multiplied by $\beta^{(\gamma_\alpha+1)K}$, is summable.

It now remains to prove (4). It suffices to show that

$$\lim_{\tau\to 0^+}\lim_{l\to\infty} \frac{M_l^\lambda}{(K_l^\lambda)^2}\mathbb{E}\left[\left(\tilde{\chi}_{n_l}^{1*}\right)^2\mathbf{1}_{\left\{\tilde{\chi}_{n_l}^{1*}\le\tau K_l^\lambda\right\}}\right] = 0. \tag{8.2}$$

Write $H_K^l(u) = \mathbb{E}\left[(\zeta_1^{l,K})^2\mathbf{1}_{\{\zeta_1^{l,K}\le u\}}\right]$ then

$$\frac{M_l^\lambda}{(K_l^\lambda)^2}\mathbb{E}\left[(\tilde{\chi}_{n_l}^{1*})^2\mathbf{1}_{\left\{\tilde{\chi}_{n_l}^{1*}\le\tau K_l^\lambda\right\}}\right] = \frac{M_l^\lambda}{(K_l^\lambda)^2}\sum_{K\in\mathbb{Z}}(c_l^K - c_l^{K+1})\beta^{2(l+K)}H_K^l(\tau\lambda\beta^{-K})$$

$$\le C_\lambda\sum_{K\in\mathbb{Z}}\beta^{(2-\gamma_\alpha)K}\beta^{\frac{\epsilon}{2}|K|}H_K^l(\tau\lambda\beta^{-K}).$$

Using that for any random variable Y we have $\mathbb{E}[Y^2 \mathbf{1}_{\{Y \leq u\}}] \leq u^\kappa \mathbb{E}[Y^{2-\kappa} \mathbf{1}_{\{Y \leq u\}}]$ with $u = \tau \lambda \beta^{-K}$ and $\kappa = 2$ it follows that

$$\sum_{K \geq 0} \beta^{(2-\gamma_\alpha)K} \beta^{\frac{\epsilon}{2}|K|} H_K^l(\tau \lambda \beta^{-K}) \leq C\tau^2 \sum_{K \geq 0} \beta^{-(\gamma_\alpha - \epsilon/2)K} \leq C\tau^2$$

where the constant C depends on $\lambda, \beta, \gamma_\alpha$ and ϵ. Then, with $u = \tau \lambda \beta^{-K}$, $\kappa = 2 - \gamma_\alpha - 2\epsilon/3$ we have that

$$H_K^l\left(\tau \lambda \beta^{-K}\right) \beta^{(2-\gamma_\alpha)K} \leq \beta^{\frac{2\epsilon}{3}K}(\tau \lambda)^{2-\gamma_\alpha - \frac{2\epsilon}{3}} \mathbb{E}\left[Z_{sup}^{\gamma_\alpha + \frac{2\epsilon}{3}}\right]$$

and therefore

$$\sum_{K \leq 0} \beta^{(2-\gamma_\alpha)K} \beta^{\frac{\epsilon}{2}|K|} H_K^l\left(\tau \lambda \beta^{-K}\right) \leq C\tau^{2-\gamma_\alpha - \frac{2\epsilon}{3}} \mathbb{E}\left[Z_{sup}^{\gamma_\alpha + \frac{2\epsilon}{3}}\right] \sum_{K \leq 0} \beta^{\frac{\epsilon}{6}K}$$

$$\leq C\tau^{2-\gamma_\alpha - \frac{2\epsilon}{3}}.$$

Since $\gamma_\alpha + \frac{2\epsilon}{3} < 1$ we have that (8.2) holds. $\square$

Combining Proposition 8.1 with Corollary 5.5 and Lemma 5.6 with

$$\lambda = \Gamma(2-\alpha)^{\frac{1}{\gamma_\alpha}} c_\mu^{\frac{1}{\gamma}} \beta^{\frac{\log(t)}{\log(\mu^{-1})} - \left\lfloor \frac{\log(t)}{\log(\mu^{-1})} \right\rfloor}$$

proves Theorem 3.

9 Tightness

We conclude the results for the walk on the subcritical tree with Theorem 4 which is a tightness result for the process and a convergence result for the scaling exponent. We only prove the result in IVIE since the proof is standard (similar to that of Theorem 1.1 of [4]) and the other cases follow by the same method; however, we state the proof more generally. Recall that r_n is a_n in IVFE, $n^{1/\gamma}$ in FVIE, $a_n^{1/\gamma}$ in IVIE and $\bar{r}_n := \max\{m \geq 0 : r_m \leq n\}$.

Proof of Theorem 4 in IVIE For statement 1 we show that $\lim_{t \to \infty} \limsup_{n \to \infty} \mathbb{P}\left(\Delta_n / r_n \notin [t^{-1}, t]\right) = 0$. Let l be such that $a_{n_l(1)} \leq a_n < a_{n_{l+1}(1)}$ then by monotonicity of Δ_n

$$\mathbb{P}\left(\frac{\Delta_n}{a_n^{1/\gamma}} \notin \left[t^{-1}, t\right]\right) \leq \mathbb{P}\left(\frac{\Delta_{n_l(1)}}{a_{n_{l+1}(1)}^{1/\gamma}} < t^{-1}\right) + \mathbb{P}\left(\frac{\Delta_{n_{l+1}(1)}}{a_{n_l(1)}^{1/\gamma}} > t\right).$$

The distribution of R_1 is continuous by Theorem III.2 of [23] since $\lim_{x\to 0} \mathcal{L}(x) = -\infty$ (where R_t denotes the limiting distribution); therefore, since the sequence $(a_{n_{l+1}(1)}/a_{n_l(1)})^{1/\gamma}$ can be bounded above by some constant c,

$$\lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\Delta_n/r_n \notin \left[t^{-1}, t\right]\right) \leq \lim_{t\to\infty} \mathbb{P}\left(R_1 \notin \left[(tc)^{-1}, tc\right]\right) = 0.$$

For statement 2 we want to show that $\lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(|X_n|/\bar{r}_n \notin [t^{-1}, t]\right) = 0$. To do this we compare $|X_n|$ with Δ_n. In order to deal with the depth X_n reaches into the traps we use a bound for the height of a trap; for any $\epsilon > 0$ we have

$$\mathbb{P}\left(\frac{|X_n|}{\bar{r}_n} \geq t\right) \leq \mathbb{P}\left(\Delta_{\lfloor t\bar{r}_n - \bar{r}_n^\epsilon \rfloor} \leq n\right) + (t\bar{r}_n - \bar{r}_n^\epsilon)\mathbb{P}\left(\mathcal{H}(\mathcal{T}^{*-}) \geq \bar{r}_n^\epsilon\right).$$

By (3.12) we have that $(t\bar{r}_n - \bar{r}_n^\epsilon)\mathbb{P}\left(\mathcal{H}(\mathcal{T}^{*-}) \geq \bar{r}_n^\epsilon\right) \to 0$ as $n \to \infty$. Using the definition of $\bar{r}_n$ we have that

$$\mathbb{P}\left(\Delta_{\lfloor t\bar{r}_n - \bar{r}_n^\epsilon \rfloor} \leq n\right) \leq \mathbb{P}\left(\frac{\Delta_{\lfloor t\bar{r}_n - \bar{r}_n^\epsilon \rfloor}}{a_{t\bar{r}_n - \bar{r}_n^\epsilon}^{1/\gamma}} \leq \frac{a_{\bar{r}_n+1}^{1/\gamma}}{a_{t\bar{r}_n - \bar{r}_n^\epsilon}^{1/\gamma}}\right).$$

Since $a_{\bar{r}_n+1}^{1/\gamma}/a_{t\bar{r}_n - \bar{r}_n^\epsilon}^{1/\gamma}$ converges to $t^{-1/\gamma\alpha}$ as $n \to \infty$, by continuity of the distribution of R_1 and statement 1 we have that $\lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(|X_n|/\bar{r}_n > t\right) = 0$.

It remains to show that $\lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(|X_n|/\bar{r}_n < t^{-1}\right) = 0$. We need to bound how far the walker backtracks after reaching a new furthest point in order to compare $|X_n|$ with Δ_n. Let $\upsilon_0 := 0$ and for $j \geq 1$ define the jth regeneration time as $\upsilon_j := \min\{m > \upsilon_{j-1} : \{X_n\}_{n=0}^{m-1} \cap \{X_n\}_{n=m}^\infty = \phi\}$ then

$$\max_{i<j\leq n} \left(|X_i| - |X_j|\right) \leq \upsilon_1 \vee \max_{2\leq i\leq n} (\upsilon_i - \upsilon_{i-1}) + \max_{0\leq i\leq n} \mathcal{H}\left(\mathcal{T}_{\rho_i}^{*-}\right).$$

The regeneration times $(\upsilon_i - \upsilon_{i-1})$, υ_1 and the heights of branches $\mathcal{H}(\mathcal{T}_{\rho_i}^{*-})$ have exponential moments for all i therefore for any $\epsilon > 0$ by a union bound

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{i<j\leq n} \left(|X_i| - |X_j|\right) > \bar{r}_n^\epsilon\right) = 0.$$

We then have that

$$\mathbb{P}\left(|X_n|/\bar{r}_n < t^{-1}\right) \leq \mathbb{P}\left(\max_{i<j\leq n} |X_i| - |X_j| > \bar{r}_n^\epsilon\right) + \mathbb{P}\left(\Delta_{\lfloor t^{-1}\bar{r}_n + \bar{r}_n^\epsilon \rfloor} > n\right)$$

$$\leq o(1) + \mathbb{P}\left(\frac{\Delta_{\lfloor 2t^{-1}\bar{r}_n \rfloor}}{a_{2t^{-1}\bar{r}_n}^{1/\gamma}} > \frac{a_{\bar{r}_n}^{1/\gamma}}{a_{2t^{-1}\bar{r}_n}^{1/\gamma}}\right).$$

Then, since $a_{\bar{r}_n}^{1/\gamma}/a_{2t^{-1}\bar{r}_n}^{1/\gamma} \to (t/2)^{1/\gamma\alpha}$ as $n \to \infty$, by continuity of the distribution of R_1 and statement 1 we indeed have that $\lim_{t\to\infty} \lim\sup_{n\to\infty} \mathbb{P}\left(|X_n|/\bar{r}_n < t^{-1}\right) = 0$.

For the final statement notice that

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{\log|X_n|}{\log(n)} \neq \gamma(\alpha-1)\right) = \mathbb{P}\left(\lim_{n\to\infty} \frac{\log|X_n|}{\log(\bar{r}_n)} \cdot \frac{\log(\bar{r}_n)}{\log(n)} \neq \gamma(\alpha-1)\right)$$

and since $\bar{r}_n = n^{\gamma(\alpha-1)}\tilde{L}(n)$ for some slowly varying function $\tilde{L}$ we have that as $n \to \infty \log(\bar{r}_n)/\log(n) \to \gamma(\alpha-1)$ thus it suffices to show that the following is equal to 0

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{\log|X_n|}{\log(\bar{r}_n)} \neq 1\right) \leq \mathbb{P}\left(\lim\sup_{n\to\infty} \frac{\log|X_n|}{\log(\bar{r}_n)} > 1\right) + \lim_{t\to\infty} \mathbb{P}\left(\lim\inf_{n\to\infty} \frac{|X_n|}{\bar{r}_n} \leq t^{-1}\right).$$

By Fatou we can bound the second term above by $\lim_{t\to\infty} \lim\inf_{n\to\infty}$ $\mathbb{P}\left(|X_n|/\bar{r}_n \leq t^{-1}\right)$ which is equal to 0 by tightness of $(|X_n|/\bar{r}_n)_{n\geq0}$.

For the first term we have

$$\mathbb{P}\left(\lim\sup_{n\to\infty} \frac{\log|X_n|}{\log(\bar{r}_n)} > 1\right) = \lim_{\varepsilon\to0^+} \mathbb{P}\left(\lim\sup_{n\to\infty} \frac{\log|X_n|}{\log(\bar{r}_n)} \geq 1+\varepsilon\right)$$

$$\leq \lim_{\varepsilon\to0^+} \mathbb{P}\left(\lim_{n\to\infty} \frac{\sup_{k\leq n}|X_n|}{\bar{r}_n^{1+\varepsilon}} \geq 1\right).$$

Writing $D'(n) := \left\{\max_{i=0,\ldots,n} \mathcal{H}(\mathcal{T}_{\rho_i}^{*-}) \leq 4\log(a_n)/\log(\mu^{-1})\right\}$ we have that $\mathbb{P}(D'(n)^c) = o(n^{-2})$ by (3.12) thus $\mathbf{P}(D'(n)^c \ i.o.) = 0$. On $D'(n)$

$$\sup_{k\leq n} |X_k| \leq |X_{\kappa_n}| + \kappa_{n+1} - \kappa_n + \frac{4\log(a_n)}{\log(\mu^{-1})}$$

where κ_n is the last regeneration time of Y before time n. Therefore, since $\kappa_{n+1} - \kappa_n$ have exponential moments we have that $\mathbb{P}(\lim\sup_{n\to\infty}(\kappa_{n+1}-\kappa_n) \geq \bar{r}_n) = 0$; hence,

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{\sup_{k\leq n}|X_n|}{\bar{r}_n^{1+\varepsilon}} \geq 1\right) \leq \mathbb{P}\left(\lim\inf_{n\to\infty} \frac{|X_{\kappa_n}|}{\bar{r}_n^{1+\varepsilon}} \geq 1 - o(1)\right)$$

$$\leq \lim_{t\to\infty} \lim\inf_{n\to\infty} \mathbb{P}\left(\frac{|X_n|}{\bar{r}_n} \geq t\right)$$

where the second inequality follows by Fatou's lemma. The result follows by tightness of $(|X_n|/\bar{r}_n)_{n\geq0}$.   □

Theorem 1 follows from Theorem 4, Proposition 6.7 and Corollary 5.6 with $\hat{\mathbb{E}}\lambda = t$ since $nq_n \sim n^\varepsilon$. More specifically, since $R_{d_t,0,\mathcal{L}_t}$ is the infinitely divisible law with characteristic exponent

$$id_1 t + \int_0^\infty e^{itx} - 1 - \frac{itx}{1+x^2} d\mathcal{L}_1(x) = \int_0^\infty e^{itx} - 1 d\mathcal{L}_1(x)$$

$$= t^{-(\alpha-1)} \int_0^\infty e^{ix} - 1 d\mathcal{L}_1(x)$$

by a simple change of variables calculation we have that the laws of the process $(\Delta_{nt}/a_n)_{t\geq 0}$ converge weakly as $n \to \infty$ under $\mathbb{P}$ with respect to the Skorohod J_1 topology on $D([0,\infty),\mathbb{R})$ to the law of the stable subordinator with characteristic function $\varphi(t) = e^{-C_\alpha t^{\alpha-1}}$ where $C_{\alpha,\beta,\mu} = -\int_0^\infty e^{ix} - 1 d\mathcal{L}_1(x)$. A straightforward calculation then shows that the Laplace transform is of the form

$$\varphi_t(s) := \mathbb{E}\left[e^{-sRF_{d_t,0,\mathcal{L}_t}}\right] = e^{-ts^{\alpha-1}C_{\alpha,\beta,\mu}}$$

where

$$C_{\alpha,\beta,\mu} = \frac{\pi(\alpha-1)}{\sin(\pi(\alpha-1))} \cdot \left(\frac{\beta(1-\beta\mu)}{2(\beta-1)}\right)^{\alpha-1}. \tag{9.1}$$

10 Supercritical tree

As discussed in the introduction, the structures of the supercritical and subcritical trees are very similar in that they consist of some backbone structure $\mathcal{Y}$ with subcritical GW-trees as leaves. The main differences are as follows:

- On the subcritical tree the backbone was a single infinite line of descent, represented by the solid line in Fig. 2 of Sect. 3. On the supercritical tree the backbone is itself a random tree, represented by the solid line in Fig. 5. In particular, it is a GW-tree without deaths whose law is determined by the generating function $g(s) := (f((1-q)s+q)-q)/(1-q)$ where f is the generating function of the original offspring law and q is the extinction probability.
- Each backbone vertex has additional children which we call buds. On the subcritical tree, the number of buds had a size-biased law independent of the position on the backbone. On the supercritical tree, the distribution over the number of buds is more complicated since it depends on the backbone. Importantly, the expected number of buds can be bounded above by $\mu(1-q)^{-1}$ independently of higher moments of the offspring law which isn't the case for the subcritical tree.
- In the subcritical case, the GW-trees forming the traps have the law of the original (unconditioned) offspring law. In the supercritical case, the law is defined by the p.g.f. $h(s) := f(qs)/q$ which has mean $f'(q)$.

Let $\mathcal{T}$ denote the supercritical tree conditioned to survive, $\mathcal{T}^\circ$ the unconditioned tree and $\mathcal{T}^-$ the tree conditioned to die out. Write Z_n, Z_n°, Z_n^- to be the size of their nth generations respectively and V_n, V_n° to be the number of vertices in the nth generation of the backbone (for $\mathcal{T}$ and $\mathcal{T}^\circ$). As in the subcritical case we denote $\mathcal{T}^{*-}$ to be a dummy branch formed by a backbone vertex, its buds and the associated traps. In Fig. 5, the dashed lines represent the finite structures comprised of the buds and leaves. It will
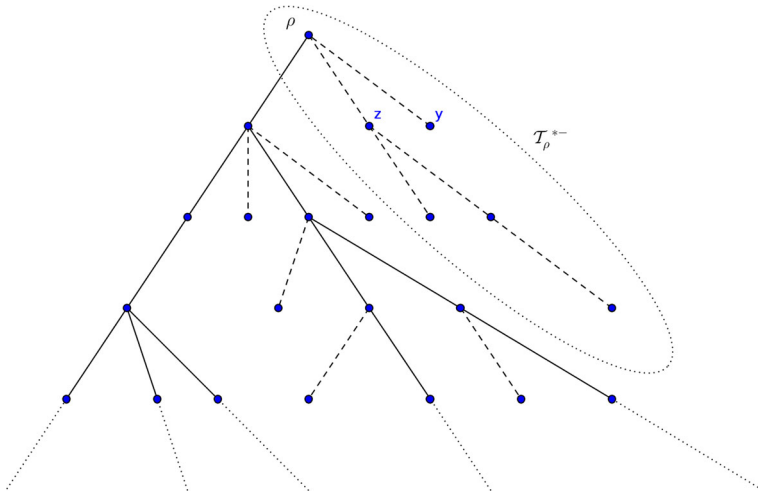
Fig. 5 A sample section of a supercritical tree with *solid lines* representing the backbone and *dashed lines* representing the dangling ends

be convenient to refer to the traps at a site so for $x \in \mathcal{Y}$ let L_x denote the collection of traps adjacent to x, for example in Fig. 5 L_ρ consists of the two trees rooted at y, z. We then write $\mathcal{T}_x^{*-}$ to be the branch at x.

Recall that Theorem 5 states that if offspring law belongs to the domain of attraction of some stable law of index $\alpha \in (1, 2)$, has mean $\mu > 1$ and the derivative of the generating function at the extinction probability satisfies $\beta > f'(q)^{-1}$. Then,

$$\frac{\Delta_{n_l(t)}}{n_l(t)^{\frac{1}{\gamma}}} \rightarrow R_t$$

in distribution as $l \to \infty$ under $\mathbb{P}$, where γ is as given in (1.1), $n_l(t) = \lfloor t f'(q)^{-l} \rfloor$ for $t > 0$ and R_t is a random variable with an infinitely divisible law whose parameters are given in [4]. Moreover, the laws of $(\Delta_n n^{-\frac{1}{\gamma}})_{n\geq 0}$ and $(|X_n| n^{-\gamma})_{n\geq 0}$ under $\mathbb{P}$ are tight on $(0, \infty)$ and $\mathbb{P}$-a.s.

$$\lim_{n\to\infty} \frac{\log |X_n|}{\log(n)} = \gamma.$$

In [4] it is shown that this holds when $\mu > 1$, $\mathbf{E}[\xi^2] < \infty$ and $\beta > f'(q)^{-1}$. In order to extend this result to prove Theorem 5 it will suffice to prove Lemmas 10.1, 10.2 and 10.3 which we defer to the end of the section.

In Lemma 10.1 we show that $\mathbf{P}(\mathcal{H}(\mathcal{T}^{*-}) > n) \sim C^* f'(q)^n$ for some constant C^*. This is the same as when $\mathbf{E}[\xi^2] < \infty$ for the supercritical tree unlike for the subcritical tree where the exponent changes depending on the stability. This is because the first moment of the bud distribution has a fundamental role and the change from finite to infinite variance changes this for the subcritical tree but not for the supercritical tree.

Lemma 10.1 is an extension of Lemma 6.1 of [4] which is proved using a Taylor expansion of the f around 1 up to second moments. We cannot take this approach because $f''(1) = \infty$; instead we use the form of the generating function determined in Lemma 3.3. The expression is important because, as in FVIE, the expected time spent in a large branch of height $\mathcal{H}(\mathcal{T}^{*-})$ is approximately $c\beta^{\mathcal{H}(\mathcal{T}^{*-})}$ for some constant c.

Lemma 10.2 shows that, with high probability, no large branch contains more than one large trap. This is important because the number of large traps would affect the escape probability. That is, if there are many large traps in a branch then it is likely that the root has many offspring on the backbone since some geometric number of the offspring lie on the backbone. The analogue of this in [4] is proved using the bound $f'(1) - f'(1-\epsilon) \leq C\epsilon$ which follows because $f''(1) < \infty$. Similarly to Lemma 10.1, we use a more precise form of f in order to obtain a similar bound.

Lemma 10.3 shows that no branch visited by level n is too large. This is important for the tightness result since we need to bound the deviation of the walk from the furthest point reached along the backbone. The proof of this follows quite straightforwardly from Lemma 10.1.

To explain why these are needed, we recall the argument which follows a similar structure to the proof of Theorem 2. As was the case for the walk on the subcritical tree, the first part of the argument involves showing that, asymptotically, the time spent outside large branches is negligible. This follows by the same techniques as for the subcritical tree.

One of the major difficulties with the walk on the supercritical tree is determining the distribution over the number of entrances into a large branch. The height of the branch from a backbone vertex x will be correlated with the number of children x has on the backbone. This affects the escape probability and therefore the number of excursions into the branch. It can be shown that the number of excursions into the first large trap converges in distribution to some non-trivial random variable W_∞. In particular, it is shown in [4] that W_∞ can be stochastically dominated by a geometric random variable and that there is some constant $c_W > 0$ such that $\mathbb{P}(W_\infty > 0) \geq c_W$.

Similarly to Sect. 4, it can be shown that asymptotically the large branches are independent in the sense that with high probability the walk won't reach one large branch and then return to a previously visited large branch. Using Lemmas 10.1 and 10.2 (among other results) it can then be shown that Δ_n can be approximated by the sum of i.i.d. random variables.

The remainder of the proof of the first part of Theorem 5 involves decomposing the time spent in large branches, showing that the suitably scaled excursion times converge in distribution, proving the convergence results for sums of i.i.d. variables and concluding with standard tightness results similar to Sect. 9. Since $\mathbf{P}(Z_1^- = k) = p_k q^{k-1}$, the subcritical GW law over the traps has exponential moments. This means that these final parts of the proof follow by the results proven in [4] since, by Lemma 10.1, the scaling is the same as when $\mathbf{E}[\xi^2] < \infty$.

Tightness of $(\Delta_n n^{-1/\gamma})_{n\geq 0}$ and $(X_n n^{-\gamma})_{n\geq 0}$ and almost sure convergence of $\log(|X_n|)/\log(n)$ then follow by the proof of Theorem 1.1 of [4] (with one slight adjustment) which is similar to the proof of Theorem 4. In order to bound the maximum distance between the walker's current position and the last regeneration point

we use a bound on the maximum height of a trap seen up to Δ_n^Y. In [4] it is shown that the probability a trap of height at least $4\log(n)/\log(f'(q)^{-1})$ is seen is at most order n^{-2} by using finite variance of the offspring distribution to bound the variance of the number of traps in a branch. In Lemma 10.3 we prove this using Lemma 10.1.

Lemma 10.1 *Under the assumptions of Theorem 5*

$$\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) > n\right) \sim C^* f'(q)^n$$

where $C^* = q(\mu - f'(q))c_\mu/(1-q)$ and c_μ is such that $\mathbf{P}(\mathcal{H}(\mathcal{T}^-) \geq n) \sim c_\mu f'(q)^n$.

Proof Let $Z := Z_1$, $Z^\circ := Z_1^\circ$, $V := V_1$, $V^\circ := V_1^\circ$ and $s_n := \mathbf{P}(\mathcal{H}(\mathcal{T}^-) < n)$, then

$$\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) > n\right) = 1 - \mathbf{E}\left[\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) \leq n|Z, V\right)\right] = 1 - \frac{\mathbf{E}\left[s_n^{Z^\circ - V^\circ}\mathbf{1}_{\{V^\circ>0\}}\right]}{1-q}.$$

For any $t, s > 0$

$$\begin{aligned}
\mathbf{E}\left[s^{Z^\circ - V^\circ}t^{V^\circ}\right] &= \mathbf{E}\left[s^{Z^\circ}\mathbf{E}[(t/s)^{V^\circ}|Z]\right] \\
&= \mathbf{E}\left[s^{Z^\circ}\sum_{k=0}^{Z^\circ}(t/s)^k\binom{Z^\circ}{k}q^{Z^\circ-k}(1-q)^k\right] \\
&= \mathbf{E}\left[(qs)^{Z^\circ}\left(1 + \frac{t(1-q)}{qs}\right)^{Z^\circ}\right] \\
&= f(sq + t(1-q)).
\end{aligned}$$

Furthermore,

$$\mathbf{E}\left[s^{Z^\circ}\mathbf{1}_{\{V^\circ=0\}}\right] = \mathbf{E}\left[s^{Z^\circ}\mathbf{P}(V^\circ = 0|Z^\circ)\right] = \mathbf{E}\left[(sq)^{Z^\circ}\right].$$

Therefore, writing $t_n := s_n q + 1 - q$ we have that $1 - t_n = q(1 - s_n)$ and

$$\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) > n\right) = 1 - \frac{f(s_n q + 1 - q)}{1-q} + \frac{f(s_n q)}{1-q} = \frac{(1 - f(t_n)) - (q - f(s_n q))}{1-q}.$$

By Taylor we have that $\exists z \in [s_n q, q]$ such that $f(s_n q) = q + qf'(q)(s_n - 1) + f''(z)q^2(s_n - 1)^2/2$. Since $q < 1$ we have that $f''(z)$ exists for all $z \leq q$ and is bounded above by $f''(q) < \infty$.

By Lemma 3.3

$$1 - f(t_n) = \mu(1 - t_n) + \frac{\Gamma(3-\alpha)}{\alpha(\alpha-1)}(1 - t_n)^\alpha \overline{L}\left((1 - t_n)^{-1}\right)$$

for a slowly varying function $\overline{L}$. In particular,

$$
\begin{aligned}
\frac{\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) > n\right)}{(1 - s_n)} &= \frac{q}{1-q}\left(\frac{1 - f(t_n)}{1 - t_n} - f'(q) + f''(z)q(s_n - 1)/2\right) \\
&= \frac{q}{1-q}\left(\mu - f'(q) + O\left((1 - s_n)^{\alpha-1}\overline{L}((1 - s_n)^{-1})\right)\right) \\
&\sim \frac{q(\mu - f'(q))}{1 - q} \tag{10.1}
\end{aligned}
$$

which is the desired result. $\qquad\square$

Recall that Δ_n^Y is the first hitting time of level n of the backbone by Y_n and L_x is the collection of traps adjacent to x then for $\varepsilon > 0$ let

$$
h_{n,\varepsilon} := \left\lceil \frac{(1 - \varepsilon)\log(n)}{\log(f'(q)^{-1})} \right\rceil \quad \text{and} \quad B(n) := \bigcap_{i=0}^{\Delta_n^Y}\{|\{T \in L_{Y_i} : \mathcal{H}(T) \geq h_{n,\varepsilon}\}| \leq 1\}
$$

denote the critical height of a trap and the event that any backbone vertex seen up to Δ_n has at most one $h_{n,\varepsilon}$-trap (which is $C_3(n)$ of [4]) respectively.

Lemma 10.2 *Under the assumptions of Theorem 5*

$$
\mathbb{P}(B(n)^c) = o(1).
$$

Proof Using $C_1(n)$ from [4] we have that the number of backbone vertices visited by Δ_n is at most Cn with high probability therefore

$$
\mathbb{P}\left(B(n)^c\right) \leq o(1) + Cn\left(\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) > h_{n,\varepsilon}\right) - \mathbf{P}(|\{T \in L_\rho : \mathcal{H}(T) \geq h_{n,\varepsilon}\}| = 1)\right).
$$

Recall $s_{h_{n,\varepsilon}} = \mathbf{P}(\mathcal{H}(\mathcal{T}^-) < h_{n,\varepsilon})$ and from (10.1) we have that

$$
\mathbf{P}\left(\mathcal{H}(\mathcal{T}^{*-}) > h_{n,\varepsilon}\right) = (1 - s_{h_{n,\varepsilon}})\frac{q\left(\mu - f'(q)\right)}{1 - q} + O((1 - s_{h_{n,\varepsilon}})^\alpha \overline{L}\left((1 - s_{h_{n,\varepsilon}})^{-1}\right)
$$

for some slowly varying function $\overline{L}$.

Similarly to the method used in Lemma 10.1 we have that

$$\mathbf{P}\left(\left|\{T \in L_\rho : \mathcal{H}(T) \geq h_{n,\varepsilon}\}\right| = 1\right)$$

$$= \sum_{k=1}^{\infty} \sum_{j=1}^{k-1} \mathbf{P}\left(Z_1 = k, V_1 = k - j, \left|\{T \in L_\rho : \mathcal{H}(T) \geq h_{n,\varepsilon}\}\right| = 1\right)$$

$$= \sum_{k=1}^{\infty} \mathbf{P}(Z_1 = k) \sum_{j=1}^{k-1} \mathbf{P}\left(V_1 = k - j \mid Z_1 = k\right) j \left(1 - s_{h_{n,\varepsilon}}\right) s_{h_{n,\varepsilon}}^{j-1}$$

$$= \sum_{k=1}^{\infty} \frac{(1 - q^k) p_k}{1 - q} \sum_{j=1}^{k-1} \binom{k}{j} \frac{q^j (1-q)^{k-j}}{1 - q^k} j \left(1 - s_{h_{n,\varepsilon}}\right) s_{h_{n,\varepsilon}}^{j-1}$$

$$= \frac{q \left(1 - s_{h_{n,\varepsilon}}\right)}{1 - q} \sum_{k=1}^{\infty} k p_k \sum_{j=0}^{k-2} \frac{(k-1)! \left(q s_{h_{n,\varepsilon}}\right)^j (1-q)^{k-1-j}}{j!(k-1-j)!}$$

$$= \frac{q \left(1 - s_{h_{n,\varepsilon}}\right)}{1 - q} \sum_{k=1}^{\infty} k p_k \left(\left(q s_{h_{n,\varepsilon}} + 1 - q\right)^{k-1} - \left(q s_{h_{n,\varepsilon}}\right)^{k-1}\right)$$

$$= \frac{q \left(1 - s_{h_{n,\varepsilon}}\right)}{1 - q} \left(f'\left(t_{h_{n,\varepsilon}}\right) - f'\left(q s_{h_{n,\varepsilon}}\right)\right)$$

where $t_{h_{n,\varepsilon}} = q s_{h_{n,\varepsilon}} + 1 - q$. By Taylor $f'(q s_n) = f'(q) + O(1 - s_{h_{n,\varepsilon}})$ as $n \to \infty$. From Lemma 3.3 we have that $1 - f(t_{h_{n,\varepsilon}}) = \mu(1 - t_{h_{n,\varepsilon}}) + (1 - t_{h_{n,\varepsilon}})^\alpha \overline{L}((1 - t_{h_{n,\varepsilon}})^{-1})$ for some slowly varying function $\overline{L}$. Applying Theorem 2 of [19] we have that $f'(t_{h_{n,\varepsilon}}) = \mu + O((1 - t_{h_{n,\varepsilon}})^{\alpha-1} \overline{L}((1 - t_{h_{n,\varepsilon}})^{-1}))$. In particular,

$$\mathbf{P}\left(\left|\{T \in L_\rho : \mathcal{H}(T) \geq h_{n,\varepsilon}\}\right| = 1\right)$$

$$= \frac{q \left(1 - s_{h_{n,\varepsilon}}\right)}{1 - q} \left(\mu - f'(q) + O\left(\left(1 - t_{h_{n,\varepsilon}}\right)^{\alpha-1} \overline{L}\left(\left(1 - t_{h_{n,\varepsilon}}\right)^{-1}\right)\right)\right)$$

since $\alpha < 2$, thus

$$\mathbb{P}(B(n)^c) \leq o(1) + O\left(n \left(1 - t_{h_{n,\varepsilon}}\right)^\alpha \overline{L}\left(\left(1 - t_{h_{n,\varepsilon}}\right)^{-1}\right)\right).$$

There exists some constant c such that $1 - t_{h_{n,\varepsilon}} \sim q c_\mu f'(q)^{h_{n,\varepsilon}} \leq c n^{-(1-\varepsilon)}$ therefore since $\alpha > 1$ we can choose $\varepsilon > 0$ small enough (depending on α) such that $\mathbb{P}(B(n)^c) = o(1)$.   □

Let $D(n) := \left\{\max_{j \leq \Delta_n^Y} \mathcal{H}(\mathcal{T}_{Y_j}^{*-}) \leq 4 \log(n)/\log(f'(q)^{-1})\right\}$ be the event that all branches seen before reaching level n are of height at most $4 \log(n)/\log(f'(q)^{-1})$.

Lemma 10.3 *Under the assumptions of Theorem 5*

$$\mathbb{P}\left(D(n)^c\right) = O(n^{-2}).$$

Proof By comparison with a biased random walk on $\mathbb{Z}$, standard large deviations estimates yield that for C sufficiently large $\mathbb{P}(\Delta_n^Y > Cn) = O(n^{-2})$. Using Lemma 10.1 we have that for independent branches $\mathcal{T}_j^{*-}$

$$\mathbf{P}\left(\bigcup_{j=1}^{C_1 n} \mathcal{H}\left(\mathcal{T}_j^{*-}\right) > \frac{4\log(n)}{\log\left(f'(q)^{-1}\right)}\right)$$

$$\leq C_1 n \mathbf{P}\left(\mathcal{H}\left(\mathcal{T}^{*-}\right) > \frac{4\log(n)}{\log\left(f'(q)^{-1}\right)}\right) \leq C n f'(q)^{\frac{4\log(n)}{\log(f'(q)^{-1})}} = Cn^{-3}.$$

$\square$

Glossary

$f(s) := \sum_{k=0}^{\infty} p_k s^k$	p.g.f. of Galton–Watson process
ξ, ξ^*	Offspring and size-biased variables
β	Bias of the walk
μ, σ^2, α	Mean, variance and stability index of the offspring distribution
γ	Scaling exponent
$T, \rho, Z_n^T, \mathcal{H}(T)$	A fixed tree, its root, nth generation size and height
$T_x, c(x), \overleftarrow{x}, d_x$	The descendent tree, children, parent and out-degree of a vertex x
$\mathcal{T}, \mathcal{Y}$	GW-tree conditioned to survive and its backbone
$\mathcal{T}^{*-}$	A dummy branch
$\mathcal{T}^\circ$	A dummy f-GW-tree
$\mathcal{T}^\leftarrow$	f-GW-tree with an artificial ancestor
$\mathcal{T}_n^\prec$	f-GW-tree conditioned to be height n
$\mathcal{T}^\prec$	An infinite trap
$\mathcal{T}^*, \overline{H}$	Pruned dummy branch and its height
$\mathcal{T}_j^+, \mathcal{T}_j^-, \rho_j^+, \rho_j^-$	Large/small traps in the pruned dummy branch and their roots
$\mathcal{T}_{i,j}^+$	jth large trap in the ith large branch
$\rho_i, (\rho_{i,j})_j$	Backbone vertex in generation i and its buds
$\delta_0, \delta_1, \ldots$	Spine of $\mathcal{T}^\prec$
δ, δ^j	Deepest vertices in $\mathcal{T}^*, \mathcal{T}_j^+$
$N(m)$	Number of traps of height at least m in $\mathcal{T}^{*-}$
N, N^i	Number of large traps in a dummy branch and ith large branch
s_m	Distribution of the height of an f-GW-tree
c_μ	Constant satisfying tail asymptotic for the height of an f-GW-tree
$\mathcal{R}^{j,k,l}$	Duration of the lth excursion from δ^j to itself on the kth excursion
X_n, Y_n	Biased random walk on $\mathcal{T}$ and the underlying walk on the backbone
η_k	Time change satisfying $Y_k = X_{\eta_k}$
Δ_n, Δ_n^Y	First hitting time of the nth level of the backbone by X_n and Y_n

υ_k	Regeneration times for the walk
$X_n^{(i)}, Y_n^{(i)};\ i \geq 1$	Independent walks on $\mathcal{T}$ and corresponding backbone walks
$\mathcal{R}_\infty$	Time taken for an excursion from the deepest point in a trap to itself
$p_1(H)$	Probability that the walk reaches δ before ρ
$p_2(H)$	Probability of escaping the tree started at δ
$P^\mathcal{T}, \mathbb{P}, \mathbf{P}$	Quenched, annealed and environment laws
$\mathbb{P}^K,\ \mathbf{P}^K$	Annealed and environment laws conditioned on the number of buds or the height
L	Slowly varying function satisfying (2.2)
a_n	Scaling sequence for ξ^* so that $\mathbf{P}(\xi^* \geq xa_n) \sim n^{-1}x^{-(\alpha-1)}$
$\tilde{L}$	Slowly varying function satisfying $a_n = n^{\frac{1}{\alpha-1}}\tilde{L}(n)$
$n_l(t)$	Subsequence for convergence
$r_n, \bar{r}_n$	Appropriate scaling of Δ_n in each case and its inverse
$l_{n,\varepsilon}, l_{n,\varepsilon}^+$	Critical number of buds in IVFE
$h_{n,\varepsilon}, h_{n,\varepsilon}^+$	Critical height of a branch in FVIE and IVIE
$\mathcal{D}^{(n)}$	Roots of large branches
$\mathcal{D}_m^{(n)}$	Roots of large branches up to level m
$K(n)$	Vertices in large traps
$(\rho_i^+)_{i\geq 1}$	Ordered roots of large branches
q_n	Probability that a backbone vertex is large
$\bar{L}_K := l_{n,0} + K$	Number of traps in a large branch
$\bar{H}_K := h_{n,0} + K$	Height of a large branch
$b_n^K := \mu^{-\bar{H}_K}/c_\mu$	Scaling for ξ^* conditioned on the height of the branch
$\chi_{t,n}$	Time spent up to $\Delta_{\lfloor nt \rfloor}$ in large traps
χ_n^i	Total time spent in large traps of the ith large branch
$\tilde{\chi}_n^i$	Time spent in the ith large branch by $X_n^{(i)}$
$\tilde{\chi}_n$	Dummy version of $\tilde{\chi}_n^i$ on $\mathcal{T}^{*-}$
$\tilde{\chi}_{t,n}$	Sum of $\tilde{\chi}_n^i$ up to $i = \lfloor ntq_n \rfloor$
$\tilde{\chi}_n^\infty$	Approximation of $\tilde{\chi}_n$
$\tilde{\chi}_n^{i*}$	Excursions from δ to δ in $\tilde{\chi}_n^i$
χ_n^*	Dummy decomposition of $\tilde{\chi}_n^{i*}$ on $\mathcal{T}^*$
$\zeta^{(n)}$	$\tilde{\chi}_n$ scaled by the number of buds or $\beta^{\bar{H}}$
Z_∞^n	Scaled excursion times
Z_∞	Limit of $Z_\infty^{n_l}$
$T^{j,k}$	The duration of the kth excursion in the jth trap of $\mathcal{T}^{*-}$
$T^{(i,j,k)}$	Duration of the kth excursion in $\mathcal{T}_{i,j}^+$
$T^{*(i,j,k)}$	Duration of the excursions from δ to δ in the kth excursion in $\mathcal{T}_{i,j}^+$
$d(x,y)$	Graph distance between points x, y
$\|x\| := d(\rho, x)$	Graph distance between x and the root of the tree
τ_x^+	First return time to vertex x

References

1. Aïdékon, E.: Speed of the biased random walk on a Galton–Watson tree. Probab. Theory Relat. Fields **159**(3–4), 597–617 (2014)
2. Athreya, K.B., Ney, P.E.: Branching Processes. Dover Publications Inc, Mineola, NY (2004)
3. Ben Arous, G., Fribergh, A.: Biased random walks on random graphs. Probab. Stat. Phys. St. Petersburg **91**, 99 (2016)
4. Ben Arous, G., Fribergh, A., Gantert, N., Hammond, A.: Biased random walks on Galton–Watson trees with leaves. Ann. Probab **40**(1), 280–338 (2012)
5. Ben Arous, G., Hammond, A.: Randomly biased walks on subcritical trees. Commun. Pure Appl. Math. **65**(11), 1481–1527 (2012)
6. Bingham, N.H., Goldie, C.M., Teugels, J.L.: Regular Variation, Volume 27 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge (1989)
7. Björnberg, J.E., Stefánsson, S.Ö.: Random walk on random infinite looptrees. J. Stat. Phys. **158**(6), 1234–1261 (2015)
8. Bowditch, A.: Biased Random Walks on Subcritical Galton–Watson Trees Conditioned to Survive. Master's thesis, University of Warwick, (2014). Available at https://www2.warwick.ac.uk/fac/sci/masdoc/people/studentpages/students2013/bowditch/
9. Bowditch, A.: Central Limit Theorems for Biased Randomly Trapped Walks on $\mathbb{Z}$. (2016). arXiv preprint arXiv:1611.06879
10. Croydon, D.A., Fribergh, A., Kumagai, T.: Biased random walk on critical Galton–Watson trees conditioned to survive. Probab. Theory Relat. Fields **157**(1–2), 453–507 (2013)
11. Dhar, D.: Diffusion and drift on percolation networks in an external field. J. Phys. **17**(5), L257 (1984)
12. Feller, W.: An Introduction to Probability Theory and its Applications, vol. II, 2nd edn. Wiley, New York (1971)
13. Fribergh, A., Hammond, A.: Phase transition for the speed of the biased random walk on the supercritical percolation cluster. Commun. Pure Appl. Math. **67**(2), 173–245 (2014)
14. Geiger, J., Kersting, G.: The Galton–Watson tree conditioned on its height. In: Probability Theory and Mathematical Statistics (Vilnius, 1998), pp. 277–286. TEV, Vilnius (1999)
15. Hammond, A.: Stable limit laws for randomly biased walks on supercritical trees. Ann. Probab. **41**(3A), 1694–1766 (2013)
16. Janson, S.: Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation. Probab. Surv. **9**, 103–252 (2012)
17. Kesten, H.: Subdiffusive behavior of random walk on a random cluster. Ann. Inst. H. Poincaré Probab. Stat. **22**(4), 425–487 (1986)
18. Kesten, H., Kozlov, M.V., Spitzer, F.: A limit law for random walk in a random environment. Compos. Math. **30**, 145–168 (1975)
19. Lamperti, J.: An occupation time theorem for a class of stochastic processes. Trans. Am. Math. Soc. **88**, 380–387 (1958)
20. Lyons, R., Pemantle, R., Peres, Y.: Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. Ann. Probab. **23**(3), 1125–1138 (1995)
21. Lyons, R., Pemantle, R., Peres, Y.: Biased random walks on Galton–Watson trees. Probab. Theory Relat. Fields **106**(2), 249–264 (1996)
22. Lyons, R., Peres, Y.: Probability on Trees and Networks. Cambridge University Press, New York (2016)
23. Petrov, V.V.: Sums of Independent Random Variables. Springer, New York (1975)
24. Sznitman, A.: On the anisotropic walk on the supercritical percolation cluster. Commun. Math. Phys. **240**(1–2), 123–148 (2003)

25. Sznitman, A., Zerner, M.: A law of large numbers for random walks in random environment. Ann. Probab. **27**(4), 1851–1869 (1999)
26. Zindy, O.: Scaling limit and aging for directed trap models. Markov Process. Relat. Fields **15**(1), 31–50 (2009)