Long paths in first passage percolation on the complete graph I. Local PWIT dynamics

Maren Eckhoff[*]   Jesse Goodman[†]   Remco van der Hofstad[‡]
Francesca R. Nardi[‡]

Abstract

We study the random geometry of first passage percolation on the complete graph equipped with independent and identically distributed edge weights. We find classes with different behaviour depending on a sequence of parameters $(s_n)_{n \geq 1}$ that quantifies the extreme-value behavior of small weights. We consider both n-independent as well as n-dependent edge weights and illustrate our results in many examples.

In particular, we investigate the case where $s_n \to \infty$, and focus on the exploration process that grows the smallest-weight tree from a vertex. We establish that the smallest-weight tree process locally converges to the invasion percolation cluster on the Poisson-weighted infinite tree, and we identify the scaling limit of the weight of the smallest-weight path between two uniform vertices. In addition, we show that over a long time interval, the growth of the smallest-weight tree maintains the same volume-height scaling exponent – volume proportional to the square of the height – found in critical Galton–Watson branching trees and critical Erdős-Rényi random graphs.

1 Model and summary of results

In this paper, we study first passage percolation on the complete graph equipped with independent and identically distributed positive and continuous edge weights. In contrast to earlier work [10, 11, 12, 15, 20], we consider the case where the extreme

---

[*]Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, United Kingdom. E-mail: eckhoff.maren@gmail.com

[†]Department of Statistics, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. E-mail: jesse.goodman@auckland.ac.nz

[‡]Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. E-mail: rhofstad@win.tue.nl, f.r.nardi@tue.nl

values of the edge weights are highly separated. We start by introducing first passage percolation (FPP). Given a graph $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$, let $(Y_e^{(\mathcal{G})})_{e \in E(\mathcal{G})}$ denote a collection of positive edge weights. Thinking of $Y_e^{(\mathcal{G})}$ as the cost of crossing an edge e, we can define a metric on $V(\mathcal{G})$ by setting

$$d_{\mathcal{G},Y^{(\mathcal{G})}}(i,j) = \inf_{\pi:\, i \to j} \sum_{e \in \pi} Y_e^{(\mathcal{G})}, \tag{1.1}$$

where the infimum is over all paths π in $\mathcal{G}$ that join i to j, and $Y^{(\mathcal{G})}$ represents the edge weights $(Y_e^{(\mathcal{G})})_{e \in E(\mathcal{G})}$. We will always assume that the infimum in (1.1) is attained uniquely, by some (finite) path $\pi_{i,j}$. We are interested in the situation where the edge weights $Y_e^{(\mathcal{G})}$ are *random*, so that $d_{\mathcal{G},Y^{(\mathcal{G})}}$ is a random metric. In particular, when the graph $\mathcal{G}$ is very large, with $|V(\mathcal{G})| = n$ say, we wish to understand the scaling behavior of the following quantities for fixed $i, j \in V(\mathcal{G})$:

(a) The *local structure* – the shape of the random neighborhood of a point;

(b) The *distance* or *total weight* $W_n = d_{\mathcal{G},Y^{(\mathcal{G})}}(i,j)$ – the total edge cost of the optimal path $\pi_{i,j}$;

(c) The *hopcount* H_n – the number of edges in the optimal path $\pi_{i,j}$.

In this paper our motivation is threefold: First, we aim to establish that the smallest-weight tree process locally converges to the invasion percolation cluster on the Poisson-weighted infinite tree (PWIT) and we identify the scaling limit of the weight of the smallest-weight path between two uniform vertices. Secondly, we aim to show that over a long time interval, the growth of the smallest-weight tree maintains the same volume-height scaling exponent found in critical Galton–Watson branching trees and critical Erdős-Rényi random graphs. Finally, we use the previous items related to problem (a), to study the FPP on the complete graph with a focus on problem (b). One could consider (a) and (b) to be *mesoscopic* or local properties, while (c) is a *macroscopic* or global property. In the companion paper [14], we will use these results to investigate the hopcount problem in (c). We will often refer to results in [14], and write, e.g., [Part II, Section 6.2] to refer to [14, Section 6.2]. We also refer to [Part II, Section 2.2] for an extended discussion of the results in these two papers and their relations to the literature.

In [10], the question was raised what the *universality classes* are for this model. We bring the discussion substantially further by describing a way to distinguish several universality classes and by identifying the limiting behavior of first passage percolation in one of these classes. The cost regime introduced in (1.1) uses the information from all edges along the path and is known as the *weak disorder* regime. By contrast, in the *strong disorder* regime the cost of a path π is given by $\max_{e \in \pi} Y_e^{(\mathcal{G})}$. We establish a firm connection between the weak and strong disorder regimes in first passage percolation. Interestingly, this connection also establishes a strong relation to invasion percolation (IP) on the PWIT, which is the scaling limit of IP on the complete graph. This process also arises as the scaling limit of the minimal spanning tree on K_n (see also [1] for the local limit of the minimal spanning tree on the complete graph).

Our main interest is in the case $\mathcal{G} = K_n$, the complete graph on n vertices $V(K_n) = [n] := \{1, \ldots, n\}$, equipped with independent and identically distributed (i.i.d.) edge weights $(Y_e^{(K_n)})_{e \in E(K_n)}$. We write Y for a random variable with $Y \stackrel{d}{=} Y_e^{(\mathcal{G})}$, and assume that the distribution function F_Y of Y is continuous. For definiteness, we study the optimal path $\pi_{1,2}$ between vertices 1 and 2; by exchangeability, $\pi_{1,2}$ has the same law as $\pi_{u,v}$ for any other $u, v \in [n]$. In [10] and [15] this setup was studied for the case that

$Y_e^{(K_n)} \stackrel{d}{=} E^s$ where E is an exponential mean 1 random variable, and $s > 0$ constant and $s = s_n > 0$ a null-sequence, respectively. First, we introduce some notation:

Notation   All limits in this paper are taken as n tends to infinity unless stated otherwise. A sequence of events $(\mathcal{A}_n)_n$ happens *with high probability (whp)* if $\mathbb{P}(\mathcal{A}_n) \to 1$. For random variables $(X_n)_n, X$, we write $X_n \stackrel{d}{\longrightarrow} X$, $X_n \stackrel{\mathbb{P}}{\longrightarrow} X$ and $X_n \stackrel{a.s.}{\longrightarrow} X$ to denote convergence in distribution, in probability and almost surely, respectively. For real-valued sequences $(a_n)_n$, $(b_n)_n$, we write $a_n = O(b_n)$ if the sequence $(a_n/b_n)_n$ is bounded; $a_n = o(b_n)$ if $a_n/b_n \to 0$; $a_n = \Theta(b_n)$ if the sequences $(a_n/b_n)_n$ and $(b_n/a_n)_n$ are both bounded; and $a_n \sim b_n$ if $a_n/b_n \to 1$. Similarly, for sequences $(X_n)_n$, $(Y_n)_n$ of random variables, we write $X_n = O_{\mathbb{P}}(Y_n)$ if the sequence $(X_n/Y_n)_n$ is tight; $X_n = o_{\mathbb{P}}(Y_n)$ if $X_n/Y_n \stackrel{\mathbb{P}}{\longrightarrow} 0$; and $X_n = \Theta_{\mathbb{P}}(Y_n)$ if the sequences $(X_n/Y_n)_n$ and $(Y_n/X_n)_n$ are both tight. Moreover, E denotes an exponentially distributed random variable with mean 1, and U denotes a random variable uniformly distributed on $[0, 1]$.

1.1   First passage percolation with heavy-tailed edge weights

In this paper, we will consider edge-weight distributions with a heavy tail near 0, in the sense that the distribution function $F_Y(y)$ decays slowly to 0 as $y \downarrow 0$. It will prove more convenient to express this notion in terms of the inverse $F_Y^{-1}(u)$ – i.e., the quantile function for Y – since we can write

$$Y_e^{(K_n)} \stackrel{d}{=} F_Y^{-1}(U), \tag{1.2}$$

where U is uniformly distributed on $[0, 1]$. Expressed in terms of F_Y^{-1}, saying that the edge-weight distribution is heavy-tailed near 0 means that $F_Y^{-1}(u)$ decays rapidly to 0 as $u \downarrow 0$. We will quantify this notion in terms of the logarithmic derivative of F_Y^{-1}, which will become large as $u \downarrow 0$.

In this section, we will assume that

$$u \frac{d}{du} \log F_Y^{-1}(u) = u^{-\alpha} L(1/u), \tag{1.3}$$

where $\alpha \geq 0$ and $t \mapsto L(t)$ is slowly varying as $t \to \infty$. In other words, we assume that $u \mapsto u\frac{d}{du} \log F_Y^{-1}(u) = \frac{d}{d(\log u)} \log F_Y^{-1}(u)$ is regularly varying as $u \downarrow 0$.

Define a sequence s_n by setting $u = 1/n$ in (1.3):

$$s_n = u \frac{d}{du} \log F_Y^{-1}(u)\Big|_{u=1/n} = \frac{(F_Y^{-1})'(1/n)}{n F_Y^{-1}(1/n)}. \tag{1.4}$$

The asymptotics of the sequence $(s_n)_n$ quantify how heavy-tailed the edge-weight distribution is. For instance, an identically constant sequence, say $s_n = s$, corresponds to a pure power law $F_Y(y) = y^{1/s}$, $F_Y^{-1}(u) = u^s$; larger values of s correspond to heavier-tailed distributions.

In this paper, we are interested in the regime where $s_n \to \infty$, which corresponds to a very heavy-tailed distribution function $F_Y(y)$ that decays to 0 slower than any power of y, as $y \downarrow 0$. Our first theorem gives the scaling of the FPP distance W_n in this regime.

Theorem 1.1 (Weight – regularly varying edge weights)**.** *Suppose that the edge-weight distribution satisfies* (1.3)*. If the sequence* $(s_n)_n$ *from* (1.4) *satisfies*

$$\frac{s_n}{\log \log n} \to \infty \qquad \text{as } n \to \infty, \tag{1.5}$$

then

$$n F_Y(W_n) \stackrel{d}{\longrightarrow} M^{(1)} \vee M^{(2)}, \tag{1.6}$$

where $M^{(1)}, M^{(2)}$ are i.i.d. random variables for which $\mathbb{P}(M^{(j)} \leq x)$ is the survival probability of a Poisson Galton–Watson branching process with mean x.

To understand the scaling in the left-hand side of (1.6), consider n i.i.d. random variables $(Y_i)_{i=1}^n$. Then

$$nF_Y \left(\min_{i=1,\dots,n} Y_i \right) \xrightarrow{d} E, \tag{1.7}$$

where E is Exponential with mean 1. We can therefore interpret the scaling in (1.6) as saying that the distance W_n lies in the same scaling regime as some of the typical values for the minimal edge weight adjacent to vertex 1. Indeed, as we shall see later, there is a single edge that makes most of the contribution to the total distance W_n. The condition $s_n/\log\log n \to \infty$ from Theorem 1.1 ensures that the contribution to W_n from all other smaller edge weights is negligible.

We now turn to the local structure. To understand the random neighbourhood of a vertex in the complete graph, we study the first passage *exploration process*. Recall from (1.1) that $d_{K_n,Y^{(K_n)}}(i,j)$ denotes the total cost of the optimal path $\pi_{i,j}$ between vertex i and j. For a vertex $i \in V(K_n)$, let the *smallest-weight tree* $\mathsf{SWT}_t^{(i)}$ be the connected subgraph of K_n defined by

$$\begin{aligned}
V(\mathsf{SWT}_t^{(i)}) &= \left\{ j \in V(K_n) \colon d_{K_n,Y^{(K_n)}}(i,j) \leq t \right\}, \\
E(\mathsf{SWT}_t^{(i)}) &= \left\{ e \in E(K_n) \colon e \in \pi_{i,j} \text{ for some } j \in V(\mathsf{SWT}_t^{(i)}) \right\}.
\end{aligned} \tag{1.8}$$

Note that $\mathsf{SWT}_t^{(i)}$ is indeed a tree: if two optimal paths $\pi_{i,k}, \pi_{i,k'}$ pass through a common vertex j, then both paths must contain $\pi_{i,j}$ since the minimizers of (1.1) are unique.

To visualize the process $(\mathsf{SWT}_t^{(i)})_{t\geq 0}$, think of the edge weight $Y_e^{(K_n)}$ as the time required for fluid to flow across the edge e. Place a source of fluid at i and allow it to spread through the graph. Then $V(\mathsf{SWT}_t^{(i)})$ is precisely the set of vertices that have been wetted by time t, while $E(\mathsf{SWT}_t^{(i)})$ is the set of edges along which, at any time up to t, fluid has flowed from a wet vertex to a previously dry vertex. Equivalently, an edge is added to $\mathsf{SWT}_t^{(i)}$ whenever it becomes completely wet, with the additional rule that an edge is not added if it would create a cycle.

In the sequel, for a subgraph $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$ of K_n, we write $\mathcal{G}$ instead of $V(\mathcal{G})$ for the vertex set when there is no risk of ambiguity. We will also write SWT_t instead of $\mathsf{SWT}_t^{(i)}$ when the choice of starting vertex i is immaterial.

To study the local structure of SWT separately from the weight W_n, we convert to discrete time by defining

$$T_{\text{size } k}^{(K_n)} = \inf \left\{ t \colon |\mathsf{SWT}_t| \geq k+1 \right\}, \qquad \mathsf{SWT}_{\text{size } k} = \mathsf{SWT}_{T_{\text{size } k}^{(K_n)}}, \tag{1.9}$$

for $k < n$. (That is, $T_{\text{size } k}^{(K_n)}$ is the time when the kth vertex, not including the starting vertex, is added to SWT.) The next theorem, which for now we state somewhat informally, says that the discrete-time FPP exploration process behaves like invasion percolation in the limit $n \to \infty$, at least in finite discrete-time windows:

Theorem 1.2. *Suppose that the edge-weight distribution satisfies* (1.3)*. If the sequence* $(s_n)_n$ *from* (1.4) *satisfies*

$$s_n \to \infty \qquad \text{as } n \to \infty \tag{1.10}$$

then, for each fixed $k_0 \in \mathbb{N}$, the sequence of trees $(\mathsf{SWT}_{\text{size } k})_{k=1}^{k_0}$, as well as the weights along its edges, converges in distribution as $n \to \infty$ to the first k_0 steps of invasion percolation on the Poisson-weighted infinite tree.

We will formalize Theorem 1.2 in Theorem 2.11 below, where we will give an explicit coupling linking the two processes. The definition of the Poisson-weighted infinite tree will be given in Section 2.2.

The next result shows that, asymptotically, SWT grows so that its size is the square of its volume, at least until it reaches diameter of order $s_n \wedge n^{1/3}$ and size of order $s_n^2 \wedge n^{2/3}$. To state the result, define $T_{\text{height } k}^{(K_n)}$ to be the first time that $\text{SWT}_t^{(j)}$ contains a vertex whose graph distance in $\text{SWT}_t^{(j)}$ from the starting vertex j, is at least k. Abbreviate (recall (1.9))

$$\text{SWT}_{\text{height } k} = \text{SWT}_{T_{\text{height } k}^{(K_n)}}. \tag{1.11}$$

Theorem 1.3. *Suppose that the edge-weight distribution satisfies (1.3). If the sequence $(s_n)_n$ from (1.4) satisfies*

$$s_n \to \infty \qquad \text{as } n \to \infty, \tag{1.12}$$

and if $(\sigma_n)_n$ is another sequence satisfying

$$\sigma_n \to \infty, \quad \sigma_n = O(s_n), \quad \sigma_n = o(n^{1/3}) \qquad \text{as } n \to \infty, \tag{1.13}$$

then

$$nF_Y\left(T_{\text{height } \sigma_n}^{(K_n)}\right) \xrightarrow{d} M^{(1)} \quad \text{and} \quad nF_Y\left(T_{\text{size } \sigma_n^2}^{(K_n)}\right) \xrightarrow{d} M^{(1)}, \tag{1.14}$$

where $\mathbb{P}(M^{(1)} \leq x)$ is the survival probability of a Poisson Galton–Watson branching process with mean x. Furthermore, $\text{SWT}_{\text{height } \sigma_n}$ contains $\Theta_{\mathbb{P}}(\sigma_n^2)$ vertices, and the diameter of $\text{SWT}_{\text{size } \sigma_n^2}$ is $\Theta_{\mathbb{P}}(\sigma_n)$.

In Theorem 1.3, the scaling of $\text{SWT}_{\text{height } \sigma_n}$ and $\text{SWT}_{\text{size } \sigma_n^2}$ – volume of the order of the square of the height – is reminiscent of critical Erdős-Rényi random graphs, or critical branching processes. This result does not apply when the spatial scale reaches heights of order $n^{1/3}$; this is the spatial scale that arises in the minimal spanning tree for the complete graph, see [2]. Likewise, the result ceases to be true when $\sigma_n \gg s_n$: in our companion paper [14], we show that a continuous-time branching process dynamics takes over for the growth of the first passage percolation smallest-weight tree. The limit in (1.14) is the analogue of (1.6) from Theorem 1.1, but does not require the assumption (1.5).

We next give some examples of edge-weight distributions that satisfy (1.3), together with their associated sequences $(s_n)_n$ defined by (1.4). In view of (1.2), if $Y \stackrel{d}{=} \tilde{g}(U)$ for an increasing function $\tilde{g} : (0,1) \to (0,\infty)$ then automatically $\tilde{g} = F_Y^{-1}$. Using the distributional identity $E \stackrel{d}{=} \log(1/U)$, it follows that if $h : (0,\infty) \to (0,\infty)$ is a decreasing function and $Y \stackrel{d}{=} h(E)$, then $F_Y^{-1}(u) = h(\log(1/u))$.

Example 1.4 (Examples of weight distributions).

(a) Let $a, \gamma > 0$. Take $Y_e^{(K_n)} \stackrel{d}{=} \exp(-aE^\gamma)$, for which $\log F_Y^{-1}(u) = -a(\log(1/u))^\gamma$ and

$$s_n = a\gamma(\log n)^{\gamma-1}. \tag{1.15}$$

The hypotheses of Theorems 1.1–1.3 are satisfied whenever $\gamma > 1$.

(b) Let $a, \gamma > 0$. Take $Y_e^{(K_n)} \stackrel{d}{=} U^{a(\log(1+\log(1/U)))^\gamma}$, for which $\log F_Y^{-1}(u) = a \log u(\log(1 + \log(1/u)))^\gamma$ and

$$s_n = a(\log(1 + \log n))^\gamma + a\gamma \frac{\log n}{1 + \log n}(\log(1 + \log n))^{\gamma-1}. \tag{1.16}$$

We note that $s_n \sim a(\log\log n)^\gamma$ as $n \to \infty$. The hypotheses of Theorems 1.2–1.3 are always satisfied, but the hypotheses of Theorem 1.1 are only satisfied for $\gamma > 1$.

(c) Let $a, \beta > 0$. Take $Y_e^{(K_n)} \stackrel{d}{=} \exp(-aU^{-\beta}/\beta)$, for which $\log F_Y^{-1}(u) = -au^{-\beta}/\beta$ and

$$s_n = an^\beta. \tag{1.17}$$

The hypotheses of Theorems 1.1–1.3 are always satisfied, but if $\beta \geq 1/3$ then we cannot apply Theorem 1.3 with $\sigma_n = s_n$.

1.2 First passage percolation with n-dependent edge weights

In Theorems 1.1–1.3, we started with a fixed edge-weight distribution and extracted a specific sequence $(s_n)_n$. For an essentially arbitrary distribution (subject to the relatively modest regular variation assumption in (1.3)), its FPP properties are fully encoded, at least for the purposes of the conclusions of Theorems 1.1–1.3, by the scaling properties of this sequence $(s_n)_n$. Thus, Theorems 1.1–1.3 show the common behaviour of a *universality class* of edge-weight distributions, and show that this universality class is described in terms of a sequence of real numbers $(s_n)_n$ and its scaling behaviour.

In this section, we reverse this setup. Indeed, we now take as input a sequence $(s_n)_n$ and consider the n-dependent edge-weight distribution

$$Y_e^{(K_n)} \stackrel{d}{=} U^{s_n}, \tag{1.18}$$

where U is uniformly distributed on $[0, 1]$. (For legibility, our notation will not indicate the implicit dependence of $Y_e^{(K_n)}$ on n.) Then the conclusions of Theorems 1.1–1.3 hold verbatim:

Theorem 1.5 (n-dependent edge weights). *Let $(s_n)_n$ be a sequence of positive numbers and consider first passage percolation with n-dependent edge weights given by* (1.18).

(a) *If $s_n/\log\log n \to \infty$ as $n \to \infty$, then*

$$nW_n^{1/s_n} \stackrel{d}{\longrightarrow} M^{(1)} \vee M^{(2)}. \tag{1.19}$$

where $M^{(1)}, M^{(2)}$ are i.i.d. random variables for which $\mathbb{P}(M^{(j)} \leq x)$ is the survival probability of a Poisson Galton–Watson branching process with mean x.

(b) *If $s_n \to \infty$ as $n \to \infty$, then for each fixed $k_0 \in \mathbb{N}$, the sequence of trees $(\mathsf{SWT}_{\mathrm{size}\,k})_{k=1}^{k_0}$, as well as the weights along its edges, converges in distribution as $n \to \infty$ to the first k_0 steps of invasion percolation on the Poisson-weighted infinite tree.*

(c) *If $s_n \to \infty$ as $n \to \infty$, and if $(\sigma_n)_n$ is another sequence satisfying $\sigma_n \to \infty$, $\sigma_n = O(s_n)$ and $\sigma_n = o(n^{1/3})$, then*

$$n\left(T_{\mathrm{height}\,\sigma_n}^{(K_n)}\right)^{1/s_n} \stackrel{d}{\longrightarrow} M^{(1)} \quad \text{and} \quad n\left(T_{\mathrm{size}\,\sigma_n^2}^{(K_n)}\right)^{1/s_n} \stackrel{d}{\longrightarrow} M^{(1)}, \tag{1.20}$$

where $\mathbb{P}(M^{(1)} \leq x)$ is the survival probability of a Poisson Galton–Watson branching process with mean x. Moreover $\mathsf{SWT}_{\mathrm{height}\,\sigma_n}$ contains $\Theta_{\mathbb{P}}(\sigma_n^2)$ vertices, and the diameter of $\mathsf{SWT}_{\mathrm{size}\,\sigma_n^2}$ is $\Theta_{\mathbb{P}}(\sigma_n)$.

The distribution function F_Y corresponding to (1.18) satisfies

$$F_Y(y) = y^{1/s_n}, \quad F_Y^{-1}(u) = u^{s_n}, \quad u\frac{d}{du}\log F_Y^{-1}(u) = s_n. \tag{1.21}$$

Thus, for each fixed n, the edge-weight distribution in (1.18) is the one for which $u\frac{d}{du}\log F_Y^{-1}(u) = s_n$ holds for all $u \in [0, 1]$ (instead of holding only when $u = 1/n$, as in (1.4) before).

To see the relationship between Theorems 1.1–1.3 and Theorem 1.5, suppose we start with a fixed edge-weight distribution Y satisfying (1.3). Define the sequence $(s_n)_n$ according to (1.4), and use that sequence as the input to Theorem 1.5. Then replacing the original edge-weight distribution by the n-dependent edge weights from (1.3) does not affect the conclusions of Theorems 1.1–1.5.

In the next section, we will explain how to prove Theorems 1.1–1.5 together by considering them in a common framework: a fixed set of underlying edge weights (which will come from the Poisson-weighted infinite tree) to which an n-dependent function is applied.

In the next section, we give a more precise setting of our results. We discuss precisely the class of weights to which our results apply, including both the U^{s_n} edge weights from (1.18), as well as the heavy-tailed edge weights in (1.2). We give an extensive outline of the proof, and formulate several results that are of independent interest.

The current paper raises the question what the precise universality classes are for first passage percolation on the complete graph. This is discussed in full detail in [Part II, Section 2.2], as it also involves the scalings of the hopcount that is not discussed further here, to which we refer the reader. A discussion of our precise results is given in Section 2.6 below.

2 Detailed results, overview and classes of edge weights

In this section, we argue that the optimal path between two vertices can be divided into two parts: The local neighbourhoods of the two endpoints that follow IP dynamics by Theorem 1.2 and 1.5 (b), and the main part of the path which is characterized in terms of a branching process. The main results of this paper connect the maximal weight $M^{(1)}$ in IP to the transition time between these two regimes, and give a detailed description of the topology of the neighbourhood contained in the IP part.

This section is organised as follows. In Section 2.1, we state the conditions on general edge weights such that our results apply. The class of such edge weights contains our key examples of E^{s_n} as well as heavy-tailed edge weights as in (1.2). In Section 2.2, we couple the first passage percolation dynamics to a continuous-time branching process, and relate it to the Poisson-weighted infinite tree (PWIT). In Section 2.3, we use this coupling to relate the first passage percolation dynamics to invasion percolation on the PWIT, which applies only to short time scales. In Section 2.4, we extend this comparison to medium time scales, and give volume estimates on the branching process up to this time scale. In Section 2.5, we explain how the first passage percolation dynamics transits from an IP dynamics up to medium time scales, to branching dynamics on the long time scale. We close this section with discussion of our detailed results in Section 2.6.

2.1 Description of the class of edge-weights to which our results apply

For fixed n, the edge weights $(Y_e^{(K_n)})_{e \in E(K_n)}$ are independent for different e. However, there is no requirement that they should be independent over n, and in Section 3, we will produce $Y_e^{(K_n)}$ using a fixed source of randomness not depending on n. Therefore, it will be useful to describe the randomness on the edge weights $((Y_e^{(K_n)})_{e \in E(K_n)} \colon n \in \mathbb{N})$ uniformly across the sequence. It will be most useful to give this description in terms of exponential random variables.

Fix independent exponential variables $(E_e^{(K_n)})_{e \in E(K_n)}$ of mean 1. We suppose that our FPP edge weights are expressed as

$$Y_e^{(K_n)} = g(E_e^{(K_n)}), \tag{2.1}$$

where $g \colon (0, \infty) \to (0, \infty)$ is a strictly increasing function. The relation between g and the distribution function F_Y is given by

$$F_Y(y) = 1 - \mathrm{e}^{-g^{-1}(y)} \quad \text{and} \quad g(x) = F_Y^{-1}\left(1 - \mathrm{e}^{-x}\right). \tag{2.2}$$

Define

$$f_n(x) = g(x/n) = F_Y^{-1}\left(1 - \mathrm{e}^{-x/n}\right). \tag{2.3}$$

Let $Y_1, \ldots, Y_n$ be i.i.d. with $Y_i = g(E_i)$ as in (2.1). Since g is increasing,

$$\min_{i \in [n]} Y_i = g\big(\min_{i \in [n]} E_i\big) \stackrel{d}{=} g(E/n) = f_n(E). \tag{2.4}$$

Because of this convenient relation between the edge weights $Y_e^{(K_n)}$ and exponential random variables, we will express our hypotheses about the distribution of the edge weights in terms of conditions on the functions $f_n(x)$ as $n \to \infty$.

From now on, we suppose that the relations (2.1)–(2.3) hold. For each n, the choice of function f_n determines the distribution of Y by $Y \stackrel{d}{=} f_n(nE)$. Since the functions $f_n(x)$ already depend explicitly on n, it will impose no extra burden to allow the distribution of Y, and the functions g and F_Y, to depend implicitly on n, although we will suppress this from the notation. Indeed, in the remainder of the paper we will refer almost exclusively to f_n rather than g or F_Y. We refer to Section 5 for a relation between the assumptions on F_Y and f_n.

We now state a general version of Theorems 1.1 and 1.5 (a), expressed in terms of the functions $f_n(x)$. For the general version of Theorems 1.2 and 1.5 (b), see Theorem 2.11 below.

Theorem 2.1. *If*

$$\lim_{n \to \infty} \frac{f_n(x + \delta)}{f_n(x) \log n} = \infty \qquad \text{for all } x \geq 1 \text{ and all } \delta > 0, \tag{2.5}$$

then

$$f_n^{-1}(W_n) \xrightarrow{d} M^{(1)} \vee M^{(2)}, \tag{2.6}$$

where $M^{(1)}, M^{(2)}$ are i.i.d. random variables for which $\mathbb{P}(M^{(j)} \leq x)$ is the survival probability of a Poisson Galton–Watson branching process with mean x.

Theorem 2.1 implies Theorem 1.1 because (2.5) follows from the assumptions of Theorem 1.1; see Lemma 5.4. Theorem 2.1 is proved in Section 4.

Theorem 2.1 gives the scaling of the FPP distance W_n in terms of f_n, describing the edge weights, as well as the random variables $M^{(1)}$ and $M^{(2)}$, describing the time it takes IP to leave an arbitrary neighborhood in two i.i.d. PWITs.

In the next theorem, we quantify the degree of heavy-tailedness of the edge-weight distribution using a sequence $(s_n)_n$. In contrast to Theorems 1.1–1.3 (where s_n is defined in terms of $f_n(\cdot)$) or Theorem 1.5 (where $f_n(\cdot)$ is defined in terms of s_n), we take as input a sequence of pairs $(f_n(\cdot), s_n)_n$ linked by the following scaling assumptions:

Condition 2.2. There exists a sequence of numbers $\xi_n < 1$ such that $s_n f_n(\xi_n) = o(f_n(1))$ and

$$\inf_{n \in \mathbb{N}} \inf_{x \in [\xi_n, C]} \frac{1}{s_n} \frac{x f_n'(x)}{f_n(x)} > 0 \qquad \text{for all } C \in (1, \infty). \tag{2.7}$$

Condition 2.3. There exists a number $\eta > 0$ such that

$$\limsup_{n \to \infty} \frac{f_n(1 + C/s_n)}{f_n(1 - \eta/s_n)} < \infty \qquad \text{for all } C \in (0, \infty). \tag{2.8}$$

The ratio $x f_n'(x)/f_n(x)$ from Condition 2.2 is the derivative of $\log f_n(x)$ with respect to $\log x$; an equivalent expression is $x g'(x)/g(x)$ with x replaced by x/n. In turn, g and F_Y^{-1} have asymptotically similar behaviour (see (2.2)) so $x f_n'(x)/f_n(x)$ is the analogue of the quantity $(F_Y^{-1})'(1/n)/n F_Y^{-1}(1/n)$ from (1.4).

In words, Condition 2.2 requires that $x f_n'(x)/f_n(x)$ grows at least as fast as order s_n, uniformly over fixed windows. Condition 2.3 can be understood as saying that, within a much smaller window, $x f_n'(x)/f_n(x)$ does not grow much faster than order s_n. In the

regularly-varying setting of Theorems 1.1–1.3, Conditions 2.2–2.3 hold with $(s_n)_n$ defined as in (1.4), see Lemma 5.4.

The next result extends Theorem 1.3 to edge weights satisfying Conditions 2.2 and 2.3:

Theorem 2.4. *Suppose $(f_n(x))_n$ and $(s_n)_n$ satisfy Conditions 2.2 and 2.3. If $(\sigma_n)_n$ is another sequence satisfying*

$$\sigma_n \to \infty, \quad \sigma_n = O(s_n), \quad \sigma_n = o(n^{1/3}) \qquad \text{as } n \to \infty, \tag{2.9}$$

then

$$f_n^{-1}\left(T_{\text{height }\sigma_n}^{(K_n)}\right) \xrightarrow{d} M^{(1)} \quad \text{and} \quad f_n^{-1}\left(T_{\text{size }\sigma_n^2}^{(K_n)}\right) \xrightarrow{d} M^{(1)}, \tag{2.10}$$

where $\mathbb{P}(M^{(1)} \leq x)$ is the survival probability of a Poisson Galton–Watson branching process with mean x. Furthermore, $\mathsf{SWT}_{\text{height }\sigma_n}$ contains $\Theta_{\mathbb{P}}(\sigma_n^2)$ vertices, and the diameter of $\mathsf{SWT}_{\text{size }\sigma_n^2}$ is $\Theta_{\mathbb{P}}(\sigma_n)$.

Theorem 2.4 is a generalization of Theorems 1.3 and 1.5 (c) because Conditions 2.2–2.3 hold in both settings: see Lemma 5.4 and the proof of these theorems in Section 7.5.

2.2 Coupling FPP to a continuous-time branching process

To study the smallest-weight tree SWT from a vertex, say vertex 1, let us consider the time until the first vertex is added. By construction, $\min_{i \in [n] \setminus \{1\}} Y_{\{1,i\}}^{(K_n)} \overset{d}{=} f_n\left(\frac{n}{n-1}E\right)$ (cf. (2.4)), where E is an exponential random variable of mean 1. We next extend this to describe the distribution of the order statistics of the weights of edges from vertex 1 to *all* other vertices.

Denote by $Y_{(k)}^{(K_n)}$ the kth smallest weight from $(Y_{\{1,i\}}^{(K_n)})_{i \in [n] \setminus \{1\}}$. Then

$$(Y_{(k)}^{(K_n)})_{k \in [n-1]} \overset{d}{=} (f_n(S_{k,n}))_{k \in [n-1]}, \tag{2.11}$$

where $S_{k,n} = \sum_{j=1}^{k} \frac{n}{n-j} E_j$ and $(E_j)_{j \in [n-1]}$ are i.i.d. exponential random variables with mean 1. The fact that the distribution of $S_{k,n}$ depends on n is awkward, and can be changed by using a *thinned* Poisson point process. Let $X_1 < X_2 < \cdots$ be the points of a Poisson point process on $(0, \infty)$ with intensity 1, so that $X_k \overset{d}{=} \sum_{j=1}^{k} E_j = \lim_{n \to \infty} S_{k,n}$. To each $k \in \mathbb{N}$, we associate a *mark* M_k which is chosen uniformly at random from $[n]$, different marks being independent. We *thin* a point X_k when $M_k = 1$ (since 1 is the initial vertex) or when $M_k = M_{k'}$ for some $k' < k$. Then

$$(Y_{(k)}^{(K_n)})_{k \in [n-1]} \overset{d}{=} (f_n(X_k))_{k \in \mathbb{N}, \, X_k \text{ unthinned}}. \tag{2.12}$$

See also [Part II, Section 3.2.1] for another way of presenting this result. In the next step, we extend this result to the smallest-weight tree SWT using a relation to FPP on the Poisson-weighted infinite tree. Before giving the definitions, we recall the Ulam–Harris notation for describing trees.

Define the tree $\mathcal{T}$ as follows. The vertices of $\mathcal{T}$ are given by finite sequences of natural numbers headed by the symbol $\varnothing_1$, which we write as $\varnothing_1 j_1 j_2 \cdots j_k$. The sequence $\varnothing_1$ denotes the root vertex of $\mathcal{T}$. We concatenate sequences $v = \varnothing_1 i_1 \cdots i_k$ and $w = \varnothing_1 j_1 \cdots j_m$ to form the sequence $vw = \varnothing_1 i_1 \cdots i_k j_1 \cdots j_m$ of length $|vw| = |v| + |w| = k + m$. Identifying a natural number j with the corresponding sequence of length 1, the jth child of a vertex v is vj, and we say that v is the parent of vj. Write $p(v)$ for the (unique) parent of $v \neq \varnothing_1$, and $p^k(v)$ for the ancestor k generations before, for $k \leq |v|$.

We can place an edge (which we could consider to be directed) between every $v \neq \varnothing_1$ and its parent; this turns $\mathcal{T}$ into a tree with root $\varnothing_1$. With a slight abuse of notation, we

will use $\mathcal{T}$ to mean both the set of vertices and the associated graph, with the edges given implicitly according to the above discussion, and we will extend this convention to any subset $\tau \subset \mathcal{T}$. We also write $\partial \tau = \{v \notin \tau \colon p(v) \in \tau\}$ for the set of children one generation away from τ.

The Poisson-weighted infinite tree is an infinite edge-weighted tree in which every vertex has infinitely many (ordered) children. To describe it formally, we associate weights to the edges of $\mathcal{T}$. By construction, we can index these edge weights by non-root vertices, writing the weights as $X = (X_v)_{v \neq \varnothing_1}$, where the weight X_v is associated to the edge between v and its parent $p(v)$. We make the convention that $X_{v0} = 0$.

Definition 2.5 (Poisson-weighted infinite tree)**.** *The* Poisson-weighted infinite tree *(PWIT) is the random tree $(\mathcal{T}, X)$ for which $X_{vk} - X_{v(k-1)}$ is exponentially distributed with mean 1, independently for each $v \in \mathcal{T}$ and each $k \in \mathbb{N}$. Equivalently, the weights $(X_{v1}, X_{v2}, \dots)$ are the (ordered) points of a Poisson point process of intensity 1 on $(0, \infty)$, independently for each v.*

Motivated by (2.12), we study FPP on $\mathcal{T}$ with edge weights $(f_n(X_v))_v$:

Definition 2.6 (First passage percolation on the Poisson-weighted infinite tree)**.** *For FPP on $\mathcal{T}$ with edge weights $(f_n(X_v))_v$, let the FPP edge weight between $v \in \mathcal{T} \setminus \{\varnothing_1\}$ and $p(v)$ be $f_n(X_v)$. The FPP distance from $\varnothing_1$ to $v \in \mathcal{T}$ is*

$$T_v = \sum_{k=0}^{|v|-1} f_n(X_{p^k(v)}) \tag{2.13}$$

and the FPP exploration process $\mathsf{BP} = (\mathsf{BP}_t)_{t \geq 0}$ on $\mathcal{T}$ is defined by $\mathsf{BP}_t = \{v \in \mathcal{T} \colon T_v \leq t\}$.

Note that the randomness in BP comes from the PWIT edge weights $(X_v)_{v \neq \varnothing_1}$, which do not depend on n. However, the FPP edge weights on $\mathcal{T}$, $T_v = f_n(X_v)$, do depend on n, and consequently the law of BP varies with n. Here and elsewhere, we will mainly suppress this n-dependence from our notation.

Note that the FPP edge weights $(f_n(X_{vk}))_{k \in \mathbb{N}}$ are themselves the points of a Poisson point process on $(0, \infty)$, independently for each $v \in \mathcal{T}$. The intensity measure of this Poisson point process, which we denote by μ_n, is the image of Lebesgue measure on $(0, \infty)$ under f_n. Since f_n is strictly increasing by assumption, μ_n has no atoms and we may abbreviate $\mu_n((a, b])$ as $\mu_n(a, b)$ for simplicity. Thus μ_n is characterized by

$$\mu_n(a, b) = f_n^{-1}(b) - f_n^{-1}(a), \qquad \int_0^\infty h(y) d\mu_n(y) = \int_0^\infty h(f_n(x)) dx, \tag{2.14}$$

for any measurable function $h \colon [0, \infty) \to [0, \infty)$.

Clearly, and as suggested by the notation, the FPP exploration process BP is a continuous-time branching process:

Proposition 2.7. *The process BP is a continuous-time branching process (CTBP), started from a single individual $\varnothing_1$, where the ages at childbearing of an individual form a Poisson point process with intensity μ_n, independently for each individual. The time T_v is the birth time $T_v = \inf \{t \geq 0 \colon v \in \mathsf{BP}_t\}$ of the individual $v \in \mathcal{T}$.*

Similar to the analysis of the weights of the edges containing vertex 1, we now introduce a thinning procedure. Define $M_{\varnothing_1} = 1$ and to each other $v \in \mathcal{T} \setminus \{\varnothing_1\}$ associate a mark M_v chosen independently and uniformly from $[n]$.

Definition 2.8 (Thinning)**.** *The vertex $v \in \mathcal{T} \setminus \{\varnothing_1\}$ is thinned if it has an ancestor $v_0 = p^k(v)$ (possibly v itself) such that $M_{v_0} = M_w$ for some unthinned vertex $w \in \mathcal{T}$ with $T_w < T_{v_0}$.*

Note that whether or not a vertex v is thinned can be assessed recursively in terms of earlier-born vertices[1] and therefore Definition 2.8 is not circular.

Write $\widetilde{\mathsf{BP}}_t$ for the subgraph of BP_t consisting of unthinned vertices. If a vertex $v \in \mathcal{T}$ is thinned, then so are all its descendants, and this implies that $\widetilde{\mathsf{BP}}_t$ is a tree for all t.

Definition 2.9. *Given a subset $\tau \subset \mathcal{T}$ and marks $M = (M_v \colon v \in \tau)$ with $M_v \in [n]$, define $\pi_M(\tau)$ to be the subgraph of K_n induced by the mapping $\tau \to [n]$, $v \mapsto M_v$. That is, $\pi_M(\tau)$ has vertex set $\{M_v \colon v \in \tau\}$, with an edge between M_v and $M_{p(v)}$ whenever $v, p(v) \in \tau$.*

Note that if the marks $(M_v)_{v \in \tau}$ are distinct then $\pi_M(\tau)$ and τ are isomorphic graphs.

The following theorem establishes a close connection between FPP on K_n and FPP on the PWIT with edge weights $(f_n(X_v))_v$:

Theorem 2.10 (Coupling to FPP on PWIT)**.** *The law of $(\mathsf{SWT}_t)_{t \geq 0}$ is the same as the law of $\left(\pi_M\big(\widetilde{\mathsf{BP}}_t\big) \right)_{t \geq 0}$.*

Theorem 2.10 is based on an explicit coupling between the edge weights $(Y_e^{(K_n)})_e$ on K_n and $(X_v)_v$ on $\mathcal{T}$. A general form of such couplings and the proof of Theorem 2.10 are given in Section 3.

2.3 Relation to invasion percolation on the PWIT: short time scales

Under our scaling assumptions, FPP on the PWIT is closely related to *invasion percolation* (IP) on the PWIT which is defined as follows. Set $\mathrm{IP}(0)$ to be the subgraph consisting of $\varnothing_1$ only. For $k \in \mathbb{N}$, form $\mathrm{IP}(k)$ inductively by adjoining to $\mathrm{IP}(k-1)$ the boundary vertex $v \in \partial \mathrm{IP}(k-1)$ connected by the boundary edge of minimal weight. We note that, since we consider only the relative ordering of the various edge weights, we can use either the PWIT edge weights $(X_v)_v$ or the FPP edge weights $(f_n(X_v))_v$.

We can now state the precise version of Theorems 1.2 and 1.5 (b).

Theorem 2.11 (Coupling to IP on the PWIT)**.** *If*

$$\lim_{n \to \infty} \frac{f_n(x + \delta)}{f_n(x)} = \infty \qquad \text{for all } x > 0 \text{ and all } \delta > 0, \tag{2.15}$$

then the smallest-weight tree SWT on K_n can be coupled to invasion percolation IP on one copy of the PWIT such that, for any fixed $k_0 \in \mathbb{N}$,

$$\mathbb{P}\big(\mathsf{SWT}_{\mathrm{size}\,k} = \pi_M(\mathrm{IP}(k)) \text{ for all } k \leq k_0\big) = 1 - o(1). \tag{2.16}$$

The convergence in Theorem 2.11 is the local weak convergence in the sense of Benjamini and Schramm [9] for appropriately chosen metrics, see also Aldous and Steele [7].

We prove Theorem 2.11 by first proving the analogous statement for the branching process approximation. Similar to (1.9), define

$$T_{\mathrm{size}\,k} = \inf \{t \colon |\mathsf{BP}_t| \geq k + 1\}, \qquad \mathsf{BP}_{\mathrm{size}\,k} = \mathsf{BP}_{T_{\mathrm{size}\,k}}. \tag{2.17}$$

Theorem 2.12 (Invasion percolation and branching processes)**.** *If (2.15) holds then, for any fixed $k_0 \in \mathbb{N}$,*

$$\mathbb{P}\big(\mathsf{BP}_{\mathrm{size}\,k} = \mathrm{IP}(k) \text{ for all } k \leq k_0\big) = 1 - o(1). \tag{2.18}$$

Theorem 2.11 follows from Theorems 2.10 and 2.12 and the observation that, whp, all of the vertices of $\mathsf{BP}_{\mathrm{size}\,k_0}$ have distinct marks and are therefore unthinned.

---

[1]At least up until the time $t = \sup_{x \geq 0} f_n(x)$ when BP_t ceases to be finite a.s. However, before time t, the root $\varnothing_1$ has had infinitely many children, a.s., so all available marks have been used and all vertices born after time t are thinned, a.s.

Write $\text{IP}(\infty) = \bigcup_{k=1}^{\infty} \text{IP}(k)$ for the limiting subgraph. We remark that $\text{IP}(\infty)$ is a strict subgraph of $\mathcal{T}$ in which every vertex has finite degree a.s. (in contrast to FPP, which eventually explores every vertex of $\mathcal{T}$).

Define
$$M^{(1)} = \sup \left\{ X_v \colon v \in \text{IP}(\infty) \setminus \{\varnothing_1\} \right\}, \tag{2.19}$$

the largest weight of an invaded edge. Then $\mathbb{P}(M^{(1)} \leq x)$ is the survival probability of a Poisson Galton–Watson branching process with mean x, as in Theorem 2.1. Indeed, define
$$\text{PGW}_x = \left\{ v \in \mathcal{T} \colon X_{p^k(v)} \leq x \text{ for all } k = 0, 1, \ldots, |v| - 1 \right\}, \tag{2.20}$$

the subset of $\mathcal{T}$ consisting of all vertices connected to the root $\varnothing_1$ by a path all of whose edge weights are x or smaller. (In other words, PGW_x is the connected component containing $\varnothing_1$ in the subgraph of $\mathcal{T}$ where we remove all edges of weight exceeding x.) Then PGW_x has the law of a Poisson Galton–Watson tree with offspring mean x, and the event $\{M^{(1)} \leq x\}$ coincides with the event $\{|\text{PGW}_x| = \infty\}$.

We remark that, a.s., the supremum in (2.19) is attained, and the unique edge of weight M is invaded after a finite number of steps. Indeed, the edge of weight $M^{(1)}$ belongs to what is called the backbone:

Definition 2.13 (The IP backbone). *The* backbone *of the IP cluster* $\text{IP}(\infty)$ *is the unique infinite oriented path in* $\text{IP}(\infty)$ *starting at the root. That is, the backbone is the unique (random) infinite sequence of vertices* $V_0^{\text{BB}}, V_1^{\text{BB}}, \ldots \in \text{IP}(\infty)$ *with* $V_0^{\text{BB}} = \varnothing_1$ *and* $p\left(V_k^{\text{BB}}\right) = V_{k-1}^{\text{BB}}$ *for all* $k \in \mathbb{N}$. *The PWIT edge weight between* V_{k-1}^{BB} *and* V_k^{BB} *is denoted* X_k^{BB}.

The statement that the edge weight $M^{(1)}$ is found along the backbone can be expressed by $M^{(1)} = \max_k X_k^{\text{BB}}$ a.s.

The value $x = 1$ acts as a *critical value* for the PWIT. Indeed, if we remove all edges of weight $X_{vk} > x$, then the subtree containing the roots is a branching process with $\text{Poi}(x)$ offspring distribution. Hence for $x \leq 1$ the tree is finite a.s., while for $x > 1$ the tree is infinite with positive probability. As a result, IP on the PWIT will have to accept edges of weight $X_{vk} > 1$ infinitely often, and we have $M^{(1)} > 1$ a.s.

For more detailed properties of the IP cluster, see Proposition 7.1 below.

2.4 Relation to invasion percolation on the PWIT: medium time scales

Theorems 2.11–2.12 state that FPP behaves like IP at short time scales. Our next two results state that this similarity is maintained over longer time scales, whose duration is given in terms of s_n.

To formalize this, recall the definition of $T_{\text{size } k}$ from (2.17), and define

$$T_{\text{height } k} = \inf \left\{ t \colon \exists v \in \text{BP}_t \text{ with } |v| \geq k \right\}. \tag{2.21}$$

We also introduce

$$T_{\text{BB } k} = \inf \left\{ t \colon V_k^{\text{BB}} \in \text{BP}_t \right\} = T_{V_k^{\text{BB}}}, \tag{2.22}$$

the first time that the FPP exploration process reaches the vertex at height k along the backbone. For $v \in \mathcal{T}^{(1)}$, write $\text{BP}^{(v)}$ for the branching process of descendants of v, re-rooted and time-shifted to start at $t = 0$. That is,

$$\text{BP}_t^{(v)} = \left\{ w \in \mathcal{T} \colon vw \in \text{BP}_{T_v + t} \right\}. \tag{2.23}$$

For instance, $\text{BP} = \text{BP}^{(\varnothing_1)}$; and the branching processes $(\text{BP}^{(v)})_{p(v) = \varnothing_1}$ are independent of each other and of $(T_v)_{p(v) = \varnothing_1}$.

Definition 2.14 (k-lucky). *Given* $k \in \mathbb{N}$, *a vertex* $v \in \mathcal{T}$ *is* k-lucky *if* $\left| \text{BP}_{f_n(1)}^{(v)} \right| \geq k$.

That is, a k-lucky vertex has k descendants by the time it reaches age $f_n(1)$. We discuss k-lucky vertices in more detail in Section 2.5, including bounds on the probability that v is k-lucky as well as the time that the first rs_n^2-lucky vertex appears for several r.

We write

$$T_{k\text{-lucky}} = f_n(1) + \inf\{T_v\colon v \text{ is } k\text{-lucky}\} \tag{2.24}$$

for the first time that a k-lucky vertex reaches age $f_n(1)$. (Thus $T_{\text{size } k} \leq T_{k\text{-lucky}}$ by definition.) Similar to (2.17), write $\mathsf{BP}_{\text{height } k}$, $\mathsf{BP}_{\mathsf{BB}\,k}$, $\mathsf{BP}_{k\text{-lucky}}$ for BP_t when $t = T_{\text{height } k}$, $t = T_{\mathsf{BB}\,k}$ or $t = T_{k\text{-lucky}}$. The analogue of Theorems 1.3, 1.5 (c) and 2.4 states that, at size scales up to order s_n^2 and distance scales up to order s_n, all these times behave similarly:

Theorem 2.15. *Suppose $(f_n(x))_n$ and $(s_n)_n$ satisfy Conditions 2.2 and 2.3. If $(\sigma_n)_n$ is another sequence satisfying*

$$\sigma_n \to \infty, \quad \sigma_n = O(s_n) \qquad \text{as } n \to \infty, \tag{2.25}$$

then

$$\left(f_n^{-1}\big(T_{\text{size } \sigma_n^2}\big), f_n^{-1}\big(T_{\text{height } \sigma_n}\big), f_n^{-1}\big(T_{\mathsf{BB}\,\sigma_n}\big), f_n^{-1}\big(T_{\sigma_n^2\text{-lucky}}\big)\right) \xrightarrow{\mathrm{P}} \left(M^{(1)}, M^{(1)}, M^{(1)}, M^{(1)}\right) \tag{2.26}$$

where $M^{(1)}$ is defined by (2.19). Moreover $\mathsf{BP}_{\text{size } \sigma_n^2}$, $\mathsf{BP}_{\text{height } \sigma_n}$, $\mathsf{BP}_{\mathsf{BB}\,\sigma_n}$ and $\mathsf{BP}_{\sigma_n^2\text{-lucky}}$ all contain $\Theta_{\mathbb{P}}(\sigma_n^2)$ vertices and have diameter $\Theta_{\mathbb{P}}(\sigma_n)$.

Note that in Theorem 2.15 we no longer need the assumption $\sigma_n = o(n^{1/3})$ from Theorem 2.4, which is only needed to ensure that vertices of $\mathsf{BP}_{\text{size } \sigma_n^2}$, $\mathsf{BP}_{\text{height } \sigma_n}$, $\mathsf{BP}_{\mathsf{BB}\,\sigma_n}$ or $\mathsf{BP}_{\sigma_n^2\text{-lucky}}$ are whp unthinned. Indeed, given a vertex v with height $O_{\mathbb{P}}(\sigma_n)$, the conditional probability that v or any of its $O_{\mathbb{P}}(\sigma_n)$ ancestors shares a mark with any of $O_{\mathbb{P}}(\sigma_n^2)$ other vertices can be upper bounded by $(1/n) \cdot O_{\mathbb{P}}(\sigma_n) \cdot O_{\mathbb{P}}(\sigma_n^2) = O_{\mathbb{P}}(\sigma_n^3/n)$, and the assumption $\sigma_n = o(n^{1/3})$ ensures that this upper bound tends to 0.

Theorem 2.15 asserts that the branching process maintains the same scaling exponents as invasion percolation – size of order the square of the height – at least until size of order s_n^2 and height of order s_n. The next result makes a stricter comparison – that FPP only explores vertices that belong to the invasion cluster – but on a somewhat shorter time scale. Define

$$T_{\text{noninvaded}} = \inf\{t\colon \mathsf{BP}_t \nsubseteq \mathrm{IP}(\infty)\}, \tag{2.27}$$

the first exploration time of a vertex that does not belong to the infinite invasion cluster. Write $\mathsf{BP}_{\text{noninvaded}}$ for BP_t with $t = T_{\text{noninvaded}}$.

Theorem 2.16. *Suppose $(f_n(x))_n$ and $(s_n)_n$ satisfy Condition 2.2. Then $|\mathsf{BP}_{\text{noninvaded}}|$ is at least of order $s_n^{4/3}$, i.e., $|\mathsf{BP}_{\text{noninvaded}}|^{-1} = O_{\mathbb{P}}(s_n^{-4/3})$.*

The proof of Theorem 2.16, which we give in Section 7.3, is based on estimating the exploration time T_v of suitably chosen vertices v. This bound is given in terms of the largest edge weight along a path leading to v, with the contribution of smaller edge weights bounded by a conditional expectation. Because of the scaling properties of the function $f_n(x)$, such expectations are overestimates of the typical size, and consequently the bound in Theorem 2.16 is not expected to be sharp.

2.5 The transition from IP to branching dynamics

Theorem 2.15 shows that the branching process BP_t, which depends implicitly on n, follows a non-standard scaling over a varying window given in terms of s_n. Eventually, however, it begins to grow in a typical branching process manner. As we shall see, this crossover occurs when k-lucky vertices, with k of order s_n^2, begin to appear (recall Definition 2.14). The following proposition investigates how likely a vertex is to be k-lucky:

Proposition 2.17 (Probability of k-luckiness). *Suppose $(f_n(x))_n$ and $(s_n)_n$ satisfy Condition 2.2. Fix $K \in (0, \infty)$. Then there exists $\delta > 0$ such that for all $k, n \in \mathbb{N}$ with $k \leq K s_n^2$,*

$$\mathbb{P}(v \text{ is } k\text{-lucky}) \geq \frac{\delta}{\sqrt{k}}. \tag{2.28}$$

Proposition 2.17 is proved in Section 6.2. We remark that the reverse inequality

$$\mathbb{P}(v \text{ is } k\text{-lucky}) \leq C/\sqrt{k} \tag{2.29}$$

is also true (and in fact holds without any assumptions on f_n; to see this, compare $\mathsf{BP}_{f_n(1)}$ to a critical Poisson Galton–Watson tree), although we will not need this result.

Taking $k = \lfloor r s_n^2 \rfloor$, we see from Proposition 2.17 that a vertex has probability of order $1/s_n$ of being $r s_n^2$-lucky. Once an $r s_n^2$-lucky vertex v is born (for some $r > 0$), another $R s_n^2$-lucky vertex (for some possibly larger $R > 0$) is likely to be born soon thereafter. Indeed, Condition 2.3 implies that between ages $f_n(1)$ and $2 f_n(1)$, the number of new children of v will be Poisson with mean of order $1/s_n$. The same is true for the order s_n^2 initial descendants of v, so that a total of order s_n children is expected to be born during this time. Among these, of order 1 can be expected to repeat the unlikely event performed by v and thereby perpetuate the growth.

To formalize this statement, fix $r > 0$, $R < \infty$ and suppose that there is a (random) vertex V such that $\mathbb{P}(V \text{ is } r s_n^2\text{-lucky}) = 1$ and $T_V + f_n(1)$ is a stopping time with respect to the filtration generated by BP_t. Let D be the set of descendants v of V satisfying $T_v \leq T_V + f_n(1)$, so that $|D| = |\mathsf{BP}_{f_n(1)}^{(V)}| \geq r s_n^2$ by definition. Let

$$\mathcal{C} = \left\{ w \colon w \text{ is } R s_n^2\text{-lucky}, \, p(w) \in D \text{ and } T_w - T_{p(w)} > f_n(1) \right\} \tag{2.30}$$

be the set of $R s_n^2$-lucky vertices born to a parent in D of age greater than $f_n(1)$. Write $V_1, V_2, \ldots$ for the vertices in $\mathcal{C}$, ordered by their birth times.

Lemma 2.18. *Suppose $(f_n(x))_n$ and $(s_n)_n$ satisfy Conditions 2.2 and 2.3. Then, with the notation above, for any fixed $k \in \mathbb{N}$ and uniformly over the random choice of vertex V,*

$$T_{V_k} - T_V = O_{\mathbb{P}}(f_n(1)). \tag{2.31}$$

Lemma 2.18 is proved in Section 6.3. It shows that, starting from time $T_{r s_n^2\text{-lucky}}$, the CTBP BP_t grows (at least) exponentially in the timescale $f_n(1)$. Indeed, we can consider each $r s_n^2$-lucky vertex V, together with its associated set D of descendants, as a composite individual. The composite children of this composite individual are the vertices $V_1, V_2, \ldots,$ together with their own associated sets of descendants. Then Lemma 2.18 states that the number of composite children born to one composite individual within time $t f_n(1)$ diverges in probability to infinity as $t \to \infty$. Because this happens uniformly in n and uniformly in the choice of composite individual, we can repeat the argument for each composite child, obtaining a supercritical CTBP of composite individuals in the timescale $f_n(1)$.

In short, once $r s_n^2$-lucky vertices begin to appear, BP_t evolves (at least) exponentially in the usual manner of a supercritical CTBP, but with composite individuals of size of order s_n^2. In [Part II, Section 3.3.1], we show (subject to mild additional assumptions about the asymptotics of $f_n(x)$ for $x \to \infty$) that this heuristic essentially finds the true rate of growth: for instance, the Malthusian parameter is asymptotically of order $f_n(1)^{-1}$, and first and second moments of $|\mathsf{BP}_t|$ scale consistently with the observation that $\mathsf{BP}_{f_n(1)}$ is either of order s_n^2 with probability of order of $1/s_n$, or else negligible if the initial vertex $\varnothing_1$ fails to be $r s_n^2$-lucky.

By contrast, before $r s_n^2$-lucky vertices appear, BP_t may take a very long time to grow. For instance, Theorem 2.15 says that, starting from a single vertex (i.e., starting

from a small fraction of a composite individual), it takes a much longer time, of order $f_n(M^{(1)}) \gg f_n(1)$, for even *one* composite individual to appear.

2.6 Discussion of our detailed results

In this section we briefly discuss our results and state open problems. For a more detailed discussion of the results in this paper and in our companion paper [14], as well as an extensive discussion of the relations to the literature, we refer to [Part II, Section 2.2].

First passage percolation (FPP) on the complete graph is closely approximated by invasion percolation (IP) on the Poisson-weighted infinite tree (PWIT), studied in [4], whenever $s_n \to \infty$. See Theorem 2.11 and the discussion in Section 2.2. However, this relationship is a *local* one, and the scaling of s_n relative to n controls whether the two objects are *globally* comparable. Theorem 2.1 shows that the weights are globally comparable provided $s_n / \log \log n \to \infty$. For the hopcount, the appropriate comparison is to the minimal spanning tree (MST) on the complete graph, obtained from running IP with a simple no-loops constraint. Path lengths in the MST scale as $n^{1/3}$ (see [3] and [2]). We conjecture that, for $s_n^3/n \to \infty$, FPP on the complete graph is in the same universality class as IP. It would be of great interest to make this connection precise when $s_n^3/n \to \infty$ by showing, for example, that $H_n/n^{1/3}$ converges in distribution, and that the scaling limit of H_n is the same as the scaling limit of the graph distance between two uniform vertices in the MST.

The local graph convergence from Theorem 2.11 and weight convergence from Theorem 2.1 are the first two in a hierarchy of possible comparisons between FPP and the MST. Strengthening the previous statement about the scaling limit of H_n, we can ask whether the optimal path between vertices $i, j \in [n]$ equals (under a suitable coupling) the unique path in the MST from i to j; whether the union of the optimal paths[2] from vertex i to every other vertex $j \neq i$ equals the entire MST; and whether these unions agree simultaneously for every $i \in [n]$. Assuming hypotheses similar to Conditions 2.2–2.3, it would be of interest to know how s_n must grow relative to n in order for each of these events to occur whp.

3 Coupling K_n and the PWIT

In Theorem 2.10, we indicated that two random processes, the first passage exploration processes $\mathsf{SWT}^{(1)}$ and $\mathsf{BP}^{(\varnothing_1)}$ on K_n and $\mathcal{T}$, respectively, could be coupled. In Section 3.1 we explain how this coupling arises as a special case of a general family of couplings between K_n, understood as a random edge-weighted graph with i.i.d. exponential edge weights, and two copies of the PWIT. In Section 3.2 we define minimal rule exploration processes and discuss the coupling in this context. In Section 3.3 and Section 3.4 we prove Theorems 2.10 and 2.11, respectively.

3.1 Exploration processes and the definition of the coupling

We consider the disjoint union $\mathcal{T}^{(1,2)} = \mathcal{T}^{(1)} \cup \mathcal{T}^{(2)}$ of two copies of the PWIT, with roots $\varnothing_1$ and $\varnothing_2$, and we assume $n \geq 2$. (In fact, the coupling we next describe works with any number of copies of the PWIT, up to a maximum of n.) As in Section 2.2, we define $M_{\varnothing_j} = j$, for $j = 1, 2$, and to each other $v \in \mathcal{T}^{(1,2)} \setminus \{\varnothing_1, \varnothing_2\}$, we associate a mark M_v chosen uniformly and independently from $[n]$. We next define an exploration process on $\mathcal{T}^{(1,2)}$:

Definition 3.1 (Exploration process on two PWITs). *Let $\mathscr{F}_0$ be a σ-field containing all*

---

[2]This union of optimal paths will be the smallest-weight tree $\mathsf{SWT}_t^{(i)}$ from Section 2 in the limit $t \to \infty$.

null sets, and let $(\mathcal{T}^{(1,2)}, X)$ be independent of $\mathscr{F}_0$. We call a sequence $\mathcal{E} = (\mathcal{E}_k)_{k \in \mathbb{N}_0}$ of subsets of $\mathcal{T}$ an exploration process *if, with probability 1, $\mathcal{E}_0 = \{\varnothing_1, \varnothing_2\}$ and, for every $k \in \mathbb{N}$, either $\mathcal{E}_k = \mathcal{E}_{k-1}$ or else $\mathcal{E}_k$ is formed by adjoining to $\mathcal{E}_{k-1}$ a previously unexplored child $v_k \in \partial \mathcal{E}_{k-1}$, where the choice of v_k depends only on the weights X_w and marks M_w for vertices $w \in \mathcal{E}_{k-1} \cup \partial \mathcal{E}_{k-1}$ and on events in $\mathscr{F}_0$.*

Examples for exploration processes are given by FPP and IP on $\mathcal{T}^{(1,2)}$. For FPP, as defined in Definition 2.6, it is necessary to convert to discrete time by observing the branching process at those moments when a new vertex is added, as in Theorem 2.12. The standard IP on $\mathcal{T}^{(1,2)}$ is defined as follows. Set $\mathrm{IP}(0) = \{\varnothing_1, \varnothing_2\}$. For $k \in \mathbb{N}$, form $\mathrm{IP}(k)$ inductively by adjoining to $\mathrm{IP}(k-1)$ the boundary vertex $v \in \partial \mathrm{IP}(k-1)$ of minimal weight. However, an exploration process is also obtained when we specify at each step (in any suitably measurable way) whether to perform an invasion step in $\mathcal{T}^{(1)}$ or in $\mathcal{T}^{(2)}$.

For $k \in \mathbb{N}$, let $\mathscr{F}_k$ be the σ-field generated by $\mathscr{F}_0$ together with the weights X_w and marks M_w for vertices $w \in \mathcal{E}_{k-1} \cup \partial \mathcal{E}_{k-1}$. Note that the requirement on the choice of v_k in Definition 3.1 can be expressed as the requirement that $\mathcal{E}$ is $(\mathscr{F}_k)_k$-adapted.

For $v \in \mathcal{T}^{(1,2)}$, define the (discrete) exploration time of v by

$$N_v = \inf \{k \in \mathbb{N}_0 : v \in \mathcal{E}_k\}. \tag{3.1}$$

Definition 3.2 (Thinning). *The vertex $v \in \mathcal{T}^{(1,2)} \setminus \{\varnothing_1, \varnothing_2\}$ is* thinned *if it has an ancestor $v_0 = p^k(v)$ (possibly v itself) such that $M_{v_0} = M_w$ for some unthinned vertex w with $N_w < N_{v_0}$. Write $\widetilde{\mathcal{E}}_k$ for the subgraph of $\mathcal{E}_k$ consisting of unthinned vertices.*

Recall the remark below Definition 2.8 that explains that the definition above is not circular.

We define the stopping times

$$N(i) = \inf \left\{k \in \mathbb{N}_0 : M_v = i \text{ for some } v \in \widetilde{\mathcal{E}}_k \right\}. \tag{3.2}$$

at which $i \in [n]$ first appears as a mark in the unthinned exploration process. Note that, on the event $\{N(i) < \infty\}$, $\widetilde{\mathcal{E}}_k$ contains a *unique* vertex in $\mathcal{T}^{(1,2)}$ whose mark is i, for any $k \geq N(i)$; call that vertex $V(i)$. On this event, we define

$$X(i, i') = \min_{w \in \mathcal{T}^{(1,2)}} \{X_w : M_w = i', p(w) = V(i)\}. \tag{3.3}$$

Lemma 3.3. *Conditional on $\mathscr{F}_{N(i)}$, and on the event $\{N(i) < \infty\}$, the distribution of $X(i, i')$ is exponential with mean n, independently for every i'. Moreover, $X(i, i')$ is $\mathscr{F}_{N(i)+1}$ measurable.*

Proof. The event $\{N(i) = k, V(i) = v\}$ is measurable with respect to the σ-field generated by $\mathscr{F}_0$ together with all edge weights X_v, X_w and marks M_v, M_w for which w is not a descendant of v. On the other hand, on $\{V(i) = v\}$, $X(i, i') = \min \{X_w : M_w = i', p(w) = v\}$ depends only on the marks and edge weights of children of v. Therefore, the distribution of $X(i, i')$ and the independence for different i' follow from the thinning property of Poisson point processes. Since the marks and edge weights of the children of $V(i)$ are measurable with respect to $\mathscr{F}_{N(i)+1}$, $X(i, i')$ is measurable with respect to this σ-field. $\square$

We define, for an edge $\{i, i'\} \in E(K_n)$,

$$X_{\{i,i'\}}^{(K_n)} = \begin{cases} \frac{1}{n} X(i, i') & \text{if } N(i) < N(i'), \\ \frac{1}{n} X(i', i) & \text{if } N(i') < N(i), \\ E_{\{i,i'\}} & \text{if } N(i) = N(i') = \infty \text{ or } N(i) = N(i') = 0, \end{cases} \tag{3.4}$$

where $(E_e)_{e \in E(K_n)}$ are exponential variables with mean 1, independent of each other and of $(X_v)_v$.

Theorem 3.4. *If $\mathcal{E}$ is an exploration process on the union $\mathcal{T}^{(1,2)}$ of two PWITs, then the edge weights $X_e^{(K_n)}$ defined in (3.4) are exponential with mean 1, independently for each $e \in E(K_n)$.*

The idea underlying Theorem 3.4 is that, by Lemma 3.3, each variable $\frac{1}{n}X(i, i')$ is exponentially distributed conditionally on the past up to the moment $N(i)$ when it may be used to set the value of $X_{\{i,i'\}}^{(K_n)}$. However, formalizing this notion requires careful attention to the relative order of the stopping times $N(i)$ and to which of the $N(i)$ are infinite.

Proof. Let $(h_e)_{e \in E(K_n)}$ be an arbitrary collection of bounded, measurable functions, and abbreviate $\langle h_e \rangle = \mathbb{E}[h_e(E)]$, where E is exponential with mean 1. It suffices to prove that

$$\mathbb{E}\Big(\prod_{e \in E(K_n)} h_e(X_e^{(K_n)}) \Big) = \prod_{e \in E(K_n)} \langle h_e \rangle. \tag{3.5}$$

We proceed by induction. To begin, we partition (3.5) according to the number $\ell \in \{0, 1, \ldots, n-2\}$ of indices $i \neq 1, 2$ for which $N(i) = \infty$, as well as the relative order of the finite values of $N(i)$. Define $i_1 = 1$, $i_2 = 2$ and, given $\ell \in [n-2]$ and $\boldsymbol{i} = (i_3, \ldots, i_{n-\ell})$, abbreviate $S_{\ell,\boldsymbol{i}} = [n] \setminus \{i_1, \ldots, i_{n-\ell}\}$.

Note that, on the event $\{N(i) = \infty \, \forall i \in S_{\ell,\boldsymbol{i}}\}$, we have $X_{\{i,j\}}^{(K_n)} = E_{\{i,j\}}$ for $\{i, j\} \subset S_{\ell,\boldsymbol{i}}$ by (3.4). The $E_{\{i,j\}}$ are exponential, independently from everything else, so we may perform the integration over these variables separately. We conclude that

$$\mathbb{E}\Big(\prod_{e \in E(K_n)} h_e(X_e^{(K_n)}) \Big)$$
$$= \sum_{(i_3, \ldots, i_n)} \mathbb{E}\Big(\mathbb{1}_{\{N(i_3) < \cdots < N(i_{n-1}) < N(i_n) \leq \infty\}} \prod_{e \in E(K_n)} h_e(X_e^{(K_n)}) \Big)$$
$$+ \sum_{\ell=2}^{n-2} \sum_{\boldsymbol{i} = (i_3, \ldots, i_{n-\ell})} \prod_{\{i,j\} \subset S_{\ell,\boldsymbol{i}}} \langle h_{\{i,j\}} \rangle \tag{3.6}$$
$$\cdot \mathbb{E}\Big(\mathbb{1}_{\{N(i_3) < \ldots < N(i_{n-\ell}) < \infty, N(i) = \infty \, \forall i \in S_{\ell,\boldsymbol{i}}\}} \prod_{\{i,j\} \not\subset S_{\ell,\boldsymbol{i}}} h_{\{i,j\}}(X_{\{i,j\}}^{(K_n)}) \Big),$$

where the first sum corresponds to $\ell = 0$ and $\ell = 1$. The sums are over vectors of distinct indices $i_3, \ldots, i_n \in \{3, \ldots, n\}$ and $i_3, \ldots, i_{n-\ell} \in \{3, \ldots, n\}$, respectively, and the notation $\{i, j\} \not\subset S_{\ell,\boldsymbol{i}}$ means that $\{i, j\}$ is an edge with at least one endpoint in $[n] \setminus S_{\ell,\boldsymbol{i}} = \{i_1, \ldots, i_{n-\ell}\}$.

In general, for $\boldsymbol{i} = (i_3, \ldots, i_{n-\ell})$ given, define the events

$$A_{\ell,\boldsymbol{i}} = \{N(i_3) < \cdots < N(i_{n-\ell}) < \infty\}, \qquad B_{\ell,\boldsymbol{i}} = \{N(i) > N(i_{n-\ell}) \, \forall i \in S_{\ell,\boldsymbol{i}}\}. \tag{3.7}$$

We claim that, for all $\ell_0 \in [n-2]$,

$$\mathbb{E}\Big(\prod_{e \in E(K_n)} h_e(X_e^{(K_n)}) \Big)$$
$$= \sum_{\boldsymbol{i} = (i_3, \ldots, i_{n-\ell_0})} \mathbb{E}\Big(\mathbb{1}_{A_{\ell_0,\boldsymbol{i}}} \mathbb{1}_{B_{\ell_0,\boldsymbol{i}}} \prod_{\{i,j\} \not\subset S_{\ell_0,\boldsymbol{i}}} h_{\{i,j\}}(X_{\{i,j\}}^{(K_n)}) \Big) \prod_{\{i,j\} \subset S_{\ell_0,\boldsymbol{i}}} \langle h_{\{i,j\}} \rangle \tag{3.8}$$
$$+ \sum_{\ell=\ell_0+1}^{n-2} \sum_{\boldsymbol{i} = (i_3, \ldots, i_{n-\ell})} \mathbb{E}\Big(\mathbb{1}_{A_{\ell,\boldsymbol{i}}} \mathbb{1}_{\{N(i) = \infty \, \forall i \in S_{\ell,\boldsymbol{i}}\}} \prod_{\{i,j\} \not\subset S_{\ell,\boldsymbol{i}}} h_{\{i,j\}}(X_{\{i,j\}}^{(K_n)}) \Big) \prod_{\{i,j\} \subset S_{\ell,\boldsymbol{i}}} \langle h_{\{i,j\}} \rangle.$$

When $\ell_0 = n - 2$, by convention, the second sum vanishes, while in the first sum $\boldsymbol{i}$ is the empty sequence.

The case $\ell_0 = 1$ reduces to (3.6): then $S_{\ell_0,\boldsymbol{i}}$ contains only one element, which we called i_n but which is in fact uniquely determined by the values $i_3, \ldots, i_{n-1}$, and the product $\prod_{\{i,j\} \subset S_{\ell_0,\boldsymbol{i}}} \langle h_{\{i,j\}} \rangle$ is empty. This initializes the induction hypothesis.

We remark that in the right-hand sides of (3.6) and (3.8), the indicators already allow us to determine which of the three cases from (3.4) occurs. For notational simplicity, we will introduce this information gradually as we proceed.

Now suppose (3.8) has been proved for a given $\ell_0 < n - 2$. In the first summand of the right-hand side of (3.8), we condition on $\mathscr{F}_{N(i_{n-\ell_0})}$. By Lemma 3.3 and the presence of the indicators, each factor $h_{\{i,j\}}(X_{\{i,j\}}^{(K_n)})$ is equal to a factor $h_{\{i,j\}}(\frac{1}{n} X(i,j))$ (or $h_{\{i,j\}}(\frac{1}{n} X(j,i))$, if $N(j) < N(i)$) that is $\mathscr{F}_{N(i_{n-\ell_0})}$-measurable, with the exception of the factors $h_{\{i_{n-\ell_0},j\}}(\frac{1}{n} X(i_{n-\ell_0}, j))$ for $j \in S_{\ell_0,\boldsymbol{i}}$, which are conditionally independent given $\mathscr{F}_{N(i_{n-\ell_0})}$, again by Lemma 3.3. Note furthermore that $A_{\ell_0,\boldsymbol{i}}, B_{\ell_0,\boldsymbol{i}} \in \mathscr{F}_{N(i_{n-\ell_0})}$, $A_{\ell_0,\boldsymbol{i}} = A_{\ell_0+1,\boldsymbol{i}} \cap \{N(i_{n-\ell_0-1}) < N(i_{n-\ell_0}) < \infty\}$ and $S_{\ell_0,\boldsymbol{i}} \cup \{i_{n-\ell_0}\} = S_{\ell_0+1,\boldsymbol{i}}$. Thus,

$$\mathbb{E}\Big(\mathbb{1}_{A_{\ell_0,\boldsymbol{i}}} \mathbb{1}_{B_{\ell_0,\boldsymbol{i}}} \prod_{\{i,j\} \not\subset S_{\ell_0,\boldsymbol{i}}} h_{\{i,j\}}(X_{\{i,j\}}^{(K_n)})\Big) \prod_{\{i,j\} \subset S_{\ell_0,\boldsymbol{i}}} \langle h_{\{i,j\}} \rangle \tag{3.9}$$

$$= \mathbb{E}\Big(\mathbb{1}_{A_{\ell_0,\boldsymbol{i}}} \mathbb{1}_{\{N(i) > N(i_{n-\ell_0}) \, \forall i \in S_{\ell_0,\boldsymbol{i}}\}} \prod_{\{i,j\} \not\subset S_{\ell_0,\boldsymbol{i}} \cup \{i_{n-\ell_0}\}} h_{\{i,j\}}(X_{\{i,j\}}^{(K_n)})\Big)$$

$$\cdot \prod_{\{i,j\} \subset S_{\ell_0,\boldsymbol{i}} \cup \{i_{n-\ell_0}\}} \langle h_{\{i,j\}} \rangle$$

$$= \mathbb{E}\Big(\mathbb{1}_{A_{\ell_0+1,\boldsymbol{i}}} \mathbb{1}_{\{N(i_{n-\ell_0-1}) < N(i_{n-\ell_0}) < N(i) \, \forall i \in S_{\ell_0,\boldsymbol{i}}\}} \prod_{\{i,j\} \not\subset S_{\ell_0+1,\boldsymbol{i}}} h_{\{i,j\}}(X_{\{i,j\}}^{(K_n)})\Big)$$

$$\cdot \prod_{\{i,j\} \subset S_{\ell_0+1,\boldsymbol{i}}} \langle h_{\{i,j\}} \rangle.$$

Leaving $i_3, \ldots, i_{n-\ell_0-1}$ fixed, we now sum (3.9) over all $i_{n-\ell_0} \in [n] \setminus \{i_1, \ldots, i_{n-\ell_0-1}\}$. This rewrites the first sum in (3.8) as

$$\sum_{\boldsymbol{i}=(i_3,\ldots,i_{n-\ell_0-1})} \mathbb{E}\Big(\mathbb{1}_{A_{\ell_0+1,\boldsymbol{i}}} \mathbb{1}_{B_{\ell_0+1,\boldsymbol{i}}} \mathbb{1}_{\{N(i) < \infty \text{ for some } i \in S_{\ell_0+1,\boldsymbol{i}}\}} \prod_{\{i,j\} \not\subset S_{\ell_0+1,\boldsymbol{i}}} h_{\{i,j\}}(X_{\{i,j\}}^{(K_n)})\Big)$$

$$\cdot \prod_{\{i,j\} \subset S_{\ell_0+1,\boldsymbol{i}}} \langle h_{\{i,j\}} \rangle. \tag{3.10}$$

However, (3.10) combines with the summand $\ell = \ell_0 + 1$ from the second sum in (3.8) (since we can rewrite $\mathbb{1}_{A_{\ell_0+1,\boldsymbol{i}}} \mathbb{1}_{\{N(i)=\infty \, \forall i \in S_{\ell_0+1,\boldsymbol{i}}\}}$ as $\mathbb{1}_{A_{\ell_0+1,\boldsymbol{i}}} \mathbb{1}_{B_{\ell_0+1,\boldsymbol{i}}} \mathbb{1}_{\{N(i)=\infty \, \forall i \in S_{\ell_0+1,\boldsymbol{i}}\}}$) to produce the first sum in (3.8) for ℓ_0 replaced by $\ell_0 + 1$. This advances the induction hypothesis, and thus completes the proof of (3.8) for all $\ell_0 \leq n - 2$.

We therefore conclude that (3.8) holds when $\ell_0 = n - 2$. In this case the second sum vanishes, while in the first sum $\boldsymbol{i}$ is the empty sequence, $S_{n-2,\boldsymbol{i}} = [n] \setminus \{1, 2\}$ and the events $A_{n-2,\boldsymbol{i}}, B_{n-2,\boldsymbol{i}}$ always occur (note that $N(i_2) = 0$). Therefore

$$\mathbb{E}\Big(\prod_{e \in E(K_n)} h_e(X_e^{(K_n)})\Big) \tag{3.11}$$

$$= \mathbb{E}\Big(\prod_{\{i,j\} \cap \{1,2\} \neq \varnothing} h_{\{i,j\}}(X_{\{i,j\}}^{(K_n)})\Big) \prod_{\{i,j\} \cap \{1,2\} = \varnothing} \langle h_{\{i,j\}} \rangle$$

$$= \mathbb{E}\Big(h_{\{1,2\}}(E_{\{1,2\}}) \prod_{i=3}^{n} h_{\{1,i\}}(\tfrac{1}{n} X(1,i)) h_{\{2,i\}}(\tfrac{1}{n} X(2,i))\Big) \prod_{\{i,j\} \cap \{1,2\} = \varnothing} \langle h_{\{i,j\}} \rangle.$$

By Lemma 3.3, $(\frac{1}{n}X(1,i))_{i\geq 3}$ and $(\frac{1}{n}X(2,i))_{i\geq 3}$ are each families of independent exponential random variables with mean 1. Moreover they are mutually independent, since they are determined from the independent Poisson point processes of edge weights corresponding to $\varnothing_1$ and $\varnothing_2$, respectively. Since furthermore $E_{\{1,2\}}$ is independent of everything, we conclude that (3.5) holds. $\qquad\square$

3.2 Minimal-rule exploration processes

An important class of exploration processes, which includes both FPP and IP, are those exploration processes determined by a minimal rule in the following sense:

Definition 3.5 (Minimal-rule exploration process). *A* minimal rule *for an exploration process $\mathcal{E}$ on $\mathcal{T}^{(1,2)}$ is an $(\mathscr{F}_k)_k$-adapted sequence $(S_k, \prec_k)_{k=1}^{\infty}$, where $S_k \subset \partial\mathcal{E}_{k-1}$ is a (possibly empty) subset of the boundary vertices of $\mathcal{E}_{k-1}$ and $\prec_k$ is a strict total ordering of the elements of S_k (if any) such that the implication*

$$w \in S_k, p(v) = p(w), M_v = M_w, X_v < X_w \implies v \in S_k, v \prec_k w \tag{3.12}$$

holds. An exploration process is determined by the minimal rule $(S_k, \prec_k)_{k=1}^{\infty}$ *if $\mathcal{E}_k = \mathcal{E}_{k-1}$ whenever $S_k = \varnothing$ and otherwise $\mathcal{E}_k$ is formed by adjoining to $\mathcal{E}_{k-1}$ the unique vertex $v_k \in S_k$ that is minimal with respect to $\prec_k$.*

In words, in every step k there is a set of boundary vertices S_k from which we can select for the next exploration step. The content of (3.12) is that, whenever a vertex $w \in S_k$ is available for selection, then all siblings of w with the same mark but smaller weight are also available for selection and are preferred over w.

For FPP on $\mathcal{T}^{(1,2)}$ with edge weights $f_n(X_v)$, we take $v \prec_k w$ if and only if $T_v < T_w$ (recall (2.13)) and take $S_k = \partial\mathcal{E}_{k-1}$. For IP on $\mathcal{T}^{(1,2)}$, we have $v \prec_k w$ if and only if $X_v < X_w$; the choice of subset S_k can be used to enforce, for instance, whether the kth step is taken in $\mathcal{T}^{(1)}$ or $\mathcal{T}^{(2)}$.

Recall the subtree $\widetilde{\mathcal{E}}_k$ of unthinned vertices from Definition 3.2 and the subgraph $\pi_M(\widetilde{\mathcal{E}}_k)$ from Definition 2.9. That is, $\pi_M(\widetilde{\mathcal{E}}_k)$ is the union of two trees with roots 1 and 2, respectively, and for $v \in \widetilde{\mathcal{E}}_k \setminus \{\varnothing_1, \varnothing_2\}$, $\pi_M(\widetilde{\mathcal{E}}_k)$ contains vertices M_v and $M_{p(v)}$ and the edge $\{M_v, M_{p(v)}\}$.

For any $i \in [n]$ for which $N(i) < \infty$, recall that $V(i)$ is the unique vertex of $\widetilde{\mathcal{E}}_k$ ($k \geq N(i)$) for which $M_{V(i)} = i$. Define $V(i, i')$ to be the first child of $V(i)$ with mark i'.

Recalling (3.3), an equivalent characterization of $V(i, i')$ is

$$X(i, i') = X_{V(i,i')}. \tag{3.13}$$

The following lemma shows that, for an exploration process determined by a minimal rule, unthinned vertices must have the form $V(i, i')$:

Lemma 3.6. *Suppose $\mathcal{E}$ is an exploration process determined by a minimal rule $(S_k, \prec_k)_{k=1}^{\infty}$ and $k \in \mathbb{N}$ is such that $\widetilde{\mathcal{E}}_k \neq \widetilde{\mathcal{E}}_{k-1}$. Let $i_k = M_{p(v_k)}$ and $i'_k = M_{v_k}$. Then $v_k = V(i_k, i'_k)$.*

Proof. By construction, $p(v_k) \in \widetilde{\mathcal{E}}_{k-1}$ and $M_{p(v_k)} = i_k$, so $V(i_k) = p(v_k)$ by definition. Moreover, $V(i_k) = p(V(i_k, i'_k))$ and $M_{V(i_k, i'_k)} = i'_k = M_{v_k}$. Suppose to the contrary that $V(i_k, i'_k) \neq v_k$. By the definition of $V(i_k, i'_k)$, it follows that $X_{V(i_k, i'_k)} < X_{v_k}$ and (3.12) yields $V(i_k, i'_k) \in S_k$ and $V(i_k, i'_k) \prec_k v_k$, a contradiction since v_k must be minimal for $\prec_k$. $\qquad\square$

If $\mathcal{E}$ is an exploration process determined by a minimal rule, then we define

$$S_k^{(K_n)} = \left\{ \{i, i'\} \in E(K_n) : i \in \pi_M(\widetilde{\mathcal{E}}_{k-1}), i' \notin \pi_M(\widetilde{\mathcal{E}}_{k-1}), V(i, i') \in S_k \right\}, \tag{3.14}$$

and

$$e_1 \mathrel{\widetilde{\prec}_k} e_2 \quad \Longleftrightarrow \quad V(i_1, i'_1) \prec_k V(i_2, i'_2), \qquad e_1, e_2 \in S_k^{(K_n)}, \tag{3.15}$$

where $e_j = \{i_j, i'_j\}$ and $i_j \in \pi_M(\widetilde{\mathcal{E}}_{k-1}), i'_j \notin \pi_M(\widetilde{\mathcal{E}}_{k-1})$ as in (3.14).

Proposition 3.7 (Thinned minimal rule). *Suppose $\mathcal{E}$ is an exploration process determined by a minimal rule $(S_k, \prec_k)_{k=1}^\infty$. Then, under the edge-weight coupling (3.4), the edge weights of $\pi_M(\widetilde{\mathcal{E}}_k)$ are determined by*

$$X^{(K_n)}_{\{M_v, M_{p(v)}\}} = \tfrac{1}{n} X_v \quad \text{for any } v \in \cup_{k=1}^\infty \widetilde{\mathcal{E}}_k \setminus \{\varnothing_1, \varnothing_2\}, \tag{3.16}$$

and generally

$$X^{(K_n)}_{\{i,i'\}} = \tfrac{1}{n} X_{V(i,i')} \quad \text{whenever} \quad i \in \pi_M(\widetilde{\mathcal{E}}_{k-1}), i' \notin \pi_M(\widetilde{\mathcal{E}}_{k-1}) \text{ for some } k \in \mathbb{N}. \tag{3.17}$$

Moreover, for any $k \in \mathbb{N}$ for which $\widetilde{\mathcal{E}}_k \neq \widetilde{\mathcal{E}}_{k-1}$, $\pi_M(\widetilde{\mathcal{E}}_k)$ is formed by adjoining to $\pi_M(\widetilde{\mathcal{E}}_{k-1})$ the unique edge $e_k \in S_k^{(K_n)}$ that is minimal with respect to $\widetilde{\prec}_k$.

Proposition 3.7 asserts that the subgraph $\pi_M(\widetilde{\mathcal{E}}_k)$ of K_n, equipped with the edge weights $(X_e^{(K_n)})_{e \in E(\pi_M(\widetilde{\mathcal{E}}_k))}$, is isomorphic as an edge-weighted graph to the subgraph $\widetilde{\mathcal{E}}_k$ of $\mathcal{T}^{(1,2)}$, equipped with the rescaled edge weights $(\tfrac{1}{n} X_v)_{v \in \widetilde{\mathcal{E}}_k \setminus \{\varnothing_1, \varnothing_2\}}$. Furthermore, the subgraphs $\pi_M(\widetilde{\mathcal{E}}_k)$ can be grown by an inductive rule. Thus the induced subgraphs $(\pi_M(\widetilde{\mathcal{E}}_k))_{k=0}^\infty$ themselves form a minimal-rule exploration process on K_n, with a minimal rule derived from that of $\mathcal{E}$, with the caveat that $\widetilde{\prec}_k$ may depend on edge weights from $\mathcal{E}_{k-1} \setminus \widetilde{\mathcal{E}}_{k-1}$ as well as from $\pi_M(\widetilde{\mathcal{E}}_{k-1})$.

Proof of Proposition 3.7. We first prove (3.17). By assumption, $N(i) \leq k - 1 < N(i')$, so (3.17) is simply the first case in (3.4) (see also (3.13)).

Take $v \in \cup_{k=1}^\infty \widetilde{\mathcal{E}}_k \setminus \{\varnothing_1, \varnothing_2\}$, and assume that $v = v_k$, i.e., set $k = N_v \geq 1$. Set $i_k = M_{p(v_k)}$ and $i'_k = M_{v_k}$. By construction, $i_k \in \pi_M(\widetilde{\mathcal{E}}_{k-1})$ but $i'_k \notin \pi_M(\widetilde{\mathcal{E}}_{k-1})$, and according to Lemma 3.6, $v_k = V(i_k, i'_k)$. So (3.16) is a special case of (3.17).

By construction, $\pi_M(\widetilde{\mathcal{E}}_k)$ is formed by adjoining to $\pi_M(\widetilde{\mathcal{E}}_{k-1})$ the vertex $i'_k = M_{v_k} \in [n]$ via the edge $e_k = \{i_k, i'_k\}$. By Lemma 3.6, $v_k = V(i_k, i'_k)$, and by the definition of a minimal rule, the vertex v_k belongs to S_k and is minimal for $\prec_k$. It follows from the definitions (3.14)–(3.15) that $e_k \in S_k^{(K_n)}$ is minimal for $\widetilde{\prec}_k$. $\qquad\square$

3.3 Coupling SWT$^{(1)}$ and BP$^{(1)}$: proof of Theorem 2.10

In this section, we prove Theorem 2.10: that is, we couple the smallest-weight tree SWT$^{(1)}$ on K_n to a single branching process BP$^{(1)}$ on $\mathcal{T}^{(1)}$. Since this statement is concerned with processes starting from only one source, we use exploration processes on $\mathcal{T}^{(1)}$ instead of $\mathcal{T}^{(1,2)}$. All results from Sections 3.1 and 3.2 carry over (indeed, the results hold for any finite number of copies of the PWIT) and the edge-weight coupling formula (3.4) becomes

$$X_e^{(K_n)} = \begin{cases} \tfrac{1}{n} X(i, i') & \text{if } N(i) < N(i'), \\ \tfrac{1}{n} X(i', i) & \text{if } N(i') < N(i), \\ E_{\{i,i'\}} & \text{if } N(i) = N(i') = \infty, \end{cases} \tag{3.18}$$

with $N(i) = N(i') = 0$ no longer possible. We are actually proving a more specific statement:

Theorem 3.8 (Coupling to FPP on PWIT). *Couple the edge weights X_v and $X_e^{(K_n)}$ according to (3.18), where the exploration process on $\mathcal{T}$ is the discrete-time FPP exploration process $(\mathrm{BP}_{\mathrm{size}\,k})_{k=0}^\infty$. Then the branching process BP and the smallest-weight tree SWT are related by*

$$\pi_M(\widetilde{\mathrm{BP}}_t) = \mathrm{SWT}_t \qquad \text{for all } t \geq 0, \text{ a.s.} \tag{3.19}$$

By Theorem 3.4, the edge weights $X_e^{(K_n)}$ are i.i.d. exponential with mean 1. Hence, as discussed in Section 2.1, the edge weights $Y_e^{(K_n)} = g(X_e^{(K_n)})$ have the distribution function F_Y. In particular, SWT has the correct marginal law under this coupling, and Theorem 3.8 contains the result of Theorem 2.10.

In the following proof we use the notation $\widetilde{\mathsf{BP}}_{\text{size } k}$ to mean $\widetilde{\mathsf{BP}}_t$ when $t = T_{\text{size } k}$, the time when BP_t first contains $k + 1$ vertices (including the starting vertex). Because of thinning, $\widetilde{\mathsf{BP}}_{\text{size } k}$ may in general have fewer than $k + 1$ vertices, and $k \mapsto \widetilde{\mathsf{BP}}_{\text{size } k}$ will not be strictly increasing.

Proof of Theorem 3.8. It is easy to verify that the discrete-time exploration process $(\mathsf{BP}_{\text{size } k})_{k=0}^{\infty}$ is determined by the minimal rule where $S_k = \partial\mathsf{BP}_{\text{size } k-1}$ and $v \prec_k^{\text{FPP}} w$ if and only if $T_v < T_w$.

The smallest-weight tree $\mathsf{SWT} = \mathsf{SWT}^{(1)}$ evolves in discrete time as follows. At time 0, SWT_0 contains only vertex 1 and no edges. The time $T_{\text{size } k'-1}^{(K_n)}$ is the time that the $(k'-1)^{\text{st}}$ vertex, not including vertex 1, is added. After time $T_{\text{size } k'-1}^{(K_n)}$, the next edge added will be the minimizer $e'_{k'}$ of $d_{K_n, Y^{(K_n)}}(1, i) + Y^{(K_n)}$ over the set of boundary edges $e = \{i, j\}$ with $i \in \mathsf{SWT}_{\text{size } k'-1}$, $j \notin \mathsf{SWT}_{\text{size } k'-1}$, and moreover $e'_{k'} = \{i'_{k'}, j'_{k'}\}$ will be added at time $T_{\text{size } k'}^{(K_n)} = d_{K_n, Y^{(K_n)}}(1, i'_{k'}) + Y_{e'_{k'}}^{(K_n)}$. It is easy to verify by induction that for any $i \in \mathsf{SWT}_{\text{size } k'}$, $d_{K_n, Y^{(K_n)}}(1, i)$ equals the sum of edge weights $Y_e^{(K_n)}$ for e belonging to the unique path in $\mathsf{SWT}_{\text{size } k'}$ from 1 to i.

Both $\widetilde{\mathsf{BP}}$ and SWT are increasing jump processes and $\pi_M(\widetilde{\mathsf{BP}}_0) = \mathsf{SWT}_0$. By an inductive argument, it suffices to show that if k, k' are such that $\widetilde{\mathsf{BP}}_{\text{size } k} \neq \widetilde{\mathsf{BP}}_{\text{size } k-1}$ and $\pi_M(\widetilde{\mathsf{BP}}_{\text{size } k-1}) = \mathsf{SWT}_{\text{size } k'-1}$ then (a) the edge $e'_{k'}$ next added to $\mathsf{SWT}_{\text{size } k'-1}$ is the same as the edge $e_k = \{i_k, i'_k\}$ that is minimal with respect to $\widetilde{\prec}_{k'}$ and therefore next added to $\pi_M(\widetilde{\mathsf{BP}}_{\text{size } k-1})$; and (b) $T_{\text{size } k'}^{(K_n)} = T_{\text{size } k}$.

Let $i \in V(\pi_M(\widetilde{\mathsf{BP}}_{\text{size } k-1}))$. The unique path in $\mathsf{SWT}_{\text{size } k'-1} = \pi_M(\widetilde{\mathsf{BP}}_{\text{size } k-1})$ from i to 1 is the image of the unique path in $\widetilde{\mathsf{BP}}_{\text{size } k-1}$ from $V(i)$ to $\varnothing_1$ under the mapping $v \mapsto M_v$ (recall Definition 2.9). According to (3.16), (2.1) and (2.3), the edge weights along this path are

$$Y_{\{M_{p^{m-1}(V(i))}, M_{p^m(V(i))}\}}^{(K_n)} = g(X_{\{M_{p^{m-1}(V(i))}, M_{p^m(V(i))}\}}^{(K_n)}) = g(\tfrac{1}{n} X_{p^{m-1}(V(i))}) = f_n(X_{p^{m-1}(V(i))}),$$
(3.20)

for $m = 1, \ldots, |V(i)|$. Summing gives $d_{K_n, Y^{(K_n)}}(1, i) = T_{V(i)}$.

In addition, let $i' \notin V(\pi_M(\widetilde{\mathsf{BP}}_{\text{size } k-1}))$ and write $e = \{i, i'\}$. By (3.17), $X_e^{(K_n)} = \tfrac{1}{n} X_{V(i,i')}$, so that $Y_e^{(K_n)} = f_n(X_{V(i,i')})$. Thus $e'_{k'}$ is the edge e that minimizes

$$d_{K_n, Y^{(K_n)}}(1, i) + Y_e^{(K_n)} = T_{V(i)} + f_n(X_{V(i,i')}) = T_{V(i,i')}$$
(3.21)

over all choices of $i' \notin V(\pi_M(\widetilde{\mathsf{BP}}_{\text{size } k-1}))$. By Proposition 3.7, so is e_k; thus (a) follows since the minimizer is unique. Moreover $T_{\text{size } k'}^{(K_n)}$ is the corresponding minimum value, namely $T_{\text{size } k'}^{(K_n)} = T_{V(i_k, i'_k)}$. According to Lemma 3.6, $v_k = V(i_k, i'_k)$ and (b) follows. □

3.4 Comparing FPP and IP: proof of Theorems 2.11 and 2.12

In this section, we prove Theorems 2.11 and 2.12 by comparing the FPP and IP dynamics on the PWIT.

Proof of Theorem 2.12. It is easy to see that IP is an exploration process determined by a minimal rule. For instance, we may take $S_k = \partial\mathsf{IP}_{k-1}$ and $v \prec_k w$ if and only if $X_v < X_w$.

In fact, it will be more convenient to use a different characterization of IP. For $v \in \mathcal{T}$, write $\mathcal{O}(v) = (X_{(v,1)}, \ldots, X_{(v,|v|)})$ for the vector of edge weights $X_{v'}$ along the

path from $\varnothing_1$ to v, ordered from largest to smallest. Set $v \prec_k^{\mathrm{IP}} w$ if and only if $\mathcal{O}(v)$ is lexicographically smaller than $\mathcal{O}(w)$. It is an elementary exercise that this minimal rule $(S_k, \prec_k^{\mathrm{IP}})_{k=1}^\infty$ also determines IP.

Let $\varepsilon > 0$ be given. Write B_{k_0} for the collection of all vertices of the form $\varnothing_1 j_1 \cdots j_r$ with $1 \le r \le k_0$ and $j_1, \ldots, j_r \le k_0$. (That is, B_{k_0} consists of all vertices in $\mathcal{T}$ within k_0 generations for which each ancestor is at most the k_0^{th} child of its parent.) Note that the first k_0 explored vertices $v_1, \ldots, v_{k_0}$ necessarily belong to B_{k_0}, for both $(\mathsf{BP}_{\mathrm{size}\,k})_{k=0}^{k_0}$ and IP. Let $\delta > 0$ and write A_δ for the event that $\inf\{X_v : v \in B_{k_0}\} \ge \delta$, $\sup\{X_v : v \in B_{k_0}\} \le 1/\delta$, and $\inf\{|X_v - X_w| : v, w \in B_{k_0}, v \ne w\} \ge \delta$. We may choose $\delta > 0$ sufficiently small that $\mathbb{P}(A_\delta) \ge 1 - \varepsilon$.

Choose $x_0 < x_1 < \cdots < x_N$ such that $x_0 = \delta$, $x_N = 1/\delta$ and $x_j - x_{j-1} \le \delta/2$ for all $j \in [N]$. By assumption, there is an $n_0 \in \mathbb{N}$ such that $f_n(x_j)/f_n(x_{j-1}) > k_0$ for all $j \in [N]$ and $n \ge n_0$. Hence, for any $x, x' \in [\delta, 1/\delta]$ with $x' \ge x + \delta$, the monotonicity of f_n implies

$$\frac{f_n(x')}{f_n(x)} \ge \frac{f_n(x_j)}{f_n(x_{j-1})} > k_0, \tag{3.22}$$

since we may choose j such that $[x_{j-1}, x_j] \subset [x, x']$. From now on assume $n \ge n_0$.

Consider any $v, w \in B_m$ such that $v \ne w$ and neither v nor w is an ancestor of the other. Then there is a smallest index j with $X_{(v,j)} \ne X_{(w,j)}$. If $X_{(v,j)} < X_{(w,j)}$ then, on A_δ,

$$\sum_{i=j}^{|v|} f_n(X_{(v,i)}) \le k_0 f_n(X_{(v,j)}) < f_n(X_{(w,j)}) \le \sum_{i=j}^{|w|} f_n(X_{(w,i)}), \tag{3.23}$$

and similarly if $X_{(v,j)} > X_{(w,j)}$. Hence $v \prec_k^{\mathrm{FPP}} w$ if and only if $\mathcal{O}(v)$ is lexicographically smaller than $\mathcal{O}(w)$, i.e. $v \prec_k^{\mathrm{IP}} w$, for any of the vertices v, w that may be relevant to $(\mathsf{BP}_{\mathrm{size}\,k})_{k=0}^{k_0}$ or $(\mathrm{IP}(k))_{k=1}^{k_0}$. Since $\mathsf{BP}_{\mathrm{size}\,0} = \mathrm{IP}(0)$, it follows that, on A_δ for n sufficiently large, we have $(\mathsf{BP}_{\mathrm{size}\,k})_{k=1}^{k_0} = (\mathrm{IP}(k))_{k=1}^{k_0}$, and since $\mathbb{P}(A_\delta) \ge 1 - \varepsilon$ with $\varepsilon > 0$ arbitrary, this completes the proof. $\qquad\square$

Proof of Theorem 2.11. Couple the edge weights on K_n according to (3.4), where the exploration process is $(\mathsf{BP}_{\mathrm{size}\,k})_{k=0}^\infty$. Fix $k_0 \in \mathbb{N}$. By Theorem 2.12, $(\mathsf{BP}_{\mathrm{size}\,k})_{k=0}^{k_0} = (\mathrm{IP}(k))_{k=1}^{k_0}$ whp. On the other hand, since k_0 is fixed, none of the first k_0 vertices explored by $(\mathsf{BP}_{\mathrm{size}\,k})_{k=0}^\infty$ is thinned, whp, so that $\pi_M(\widetilde{\mathsf{BP}}_{\mathrm{size}\,k}) = \mathsf{SWT}_{\mathrm{size}\,k}$ whp by Theorem 2.10. $\qquad\square$

4 A strong disorder result: proof of Theorem 2.1

Proof of Theorem 2.1. The proof of (2.6) proceeds via stochastic upper and lower bounds on W_n based on couplings with IP where we use two different exploration processes. For the lower bound, consider the minimal-rule exploration process $\mathcal{E} = (\mathcal{E}_k)_{k \in \mathbb{N}_0}$ given by IP on $\mathcal{T}^{(1,2)}$ that alternates between an invasion step in $\mathcal{T}^{(1)}$ and an invasion step in $\mathcal{T}^{(2)}$ as explained below Definition 3.5. By Theorem 3.4 the edge weights $Y_e^{(K_n)} = f_n(X_e^{(K_n)})$ derived from this exploration process as in (3.4) are i.i.d. with distribution function F_Y and it suffices to consider the FPP problem on K_n with these edge weights. Let $\mathscr{V}^{(j)}$ denote the set of vertices in the invasion cluster on $\mathcal{T}^{(j)}$ explored before the edge of weight $M^{(j)}$ is invaded. (In the language of [4, 13, 17, 21], $\mathscr{V}^{(j)}$ is the first pond, not including the first outlet.) Let $\mathscr{V}_+^{(j)}$ consist of $\mathscr{V}^{(j)}$ together with all adjacent vertices connected by an edge of weight at most $M^{(1)} \vee M^{(2)}$.

Let $\mathcal{A}_n$ be the event that none of the vertices in $\mathscr{V}_+^{(1)} \cup \mathscr{V}_+^{(2)}$ are thinned, and that the exponential variable $X_{\{1,2\}}^{(K_n)} = E_{\{1,2\}}$ from (3.4) satisfies $X_{\{1,2\}}^{(K_n)} \ge \frac{1}{n}(M^{(1)} \vee M^{(2)})$. Since $\mathscr{V}_+^{(j)}$ is finite and the vertices in $\mathscr{V}^{(j)}$ are explored after a finite number of steps (not depending on n), $\mathcal{A}_n$ holds with high probability.

On $\mathcal{A}_n$, let $\mathscr{W}^{(j)}$ denote the image in $[n]$ of $\mathscr{V}^{(j)}$ under the mapping $\pi_M \colon v \mapsto M_v$ from Definition 2.9. Then every edge e between $\mathscr{W}^{(j)}$ and $[n] \setminus (\mathscr{W}^{(1)} \cup \mathscr{W}^{(2)})$ satisfies $Y_e^{(K_n)} \geq f_n(M^{(j)})$, and every edge e between $\mathscr{W}^{(1)}$ and $\mathscr{W}^{(2)}$ satisfies $Y_e^{(K_n)} \geq f_n(M^{(1)} \vee M^{(2)})$. Since every path between vertices 1 and 2 has to leave $\mathscr{W}^{(1)}$ and $\mathscr{W}^{(2)}$, this therefore proves that $W_n \geq f_n(M^{(1)} \vee M^{(2)})$ on $\mathcal{A}_n$, i.e., whp.

For the upper bound, let $\varepsilon > 0$ and let $N \in \mathbb{N}$ denote a constant to be chosen later. Modify the minimal-rule exploration process above by stopping after N steps in each subtree $\mathcal{T}^{(j)}$, i.e., set $\mathcal{E}'_k = \mathcal{E}_{k \wedge 2N}$, and couple the edge weights according to (3.4). Denote by $X_{\max}^{(j)}$ the largest edge weight in $\mathcal{T}^{(j)}$ so invaded, so that $X_{\max}^{(j)} \leq M^{(j)}$ by definition.

Let $\mathscr{U}^{(j)} = \{v \in \partial \mathcal{E}_{2N_1} \cap \mathcal{T}^{(j)} \colon X_{\max}^{(j)} < X_v < X_{\max}^{(j)} + \varepsilon\}$ denote the collection of boundary vertices joined to invaded vertices by an edge of weight at most $X_{\max}^{(j)} + \varepsilon$. Conditional on $X_{\max}^{(1)}, X_{\max}^{(2)}$ and $\mathcal{E}'$, the number $|\mathscr{U}^{(j)}|$ of such boundary vertices is Poisson with mean εN, independently for $j \in \{1, 2\}$. (This holds because the event that the exploration process $\mathcal{E}'$ explores a given sequence of vertices $v_1, \ldots, v_{2N_1}$ can be expressed solely in terms of the numbers $|\{vw \in \partial \mathcal{E}'_k \colon X_{vw} < X_{v_i}\}|$ of boundary edges of *smaller* weight, over all $k, i = 1, \ldots, 2N$.)

Let $\mathcal{A}'_n$ denote the event that none of the vertices in $\mathcal{E}'_{2N} \cup \mathscr{U}^{(1)} \cup \mathscr{U}^{(2)}$ have the same mark. Condition on the occurrence of $\mathcal{A}'_n$ and on the values of the disjoint vertex sets $\pi_M(\mathcal{E}'_{2N}), \pi_M(\mathscr{U}^{(1)}), \pi_M(\mathscr{U}^{(2)})$. Consider the induced subgraph K'_{n-2N-2} of K_n obtained by excluding the $2N + 2$ explored vertices in $\pi_M(\mathcal{E}'_{2N})$. Since no other vertices are explored, the edge weights in this induced subgraph are the independent exponential random variables E_e from (3.4).

The random subgraph $G'_{n-2N-2} = \{e \in E(K'_{n-2N-2}) \colon E_e \leq \frac{1}{n}(1 + \varepsilon)\}$ has the (conditional) law of the Erdős-Rényi random graph $G(n - 2N - 2, p)$ with $p = \mathbb{P}(E_e \leq \frac{1}{n}(1 + \varepsilon)) \sim \frac{1}{n}(1 + \varepsilon)$ as $n \to \infty$. As is well known, in the supercritical regime, the giant component has diameter $O_{\mathbb{P}}(\log n)$ and contains a positive asymptotic fraction of vertices (see e.g., [16]). Suppose $U^{(1)}, U^{(2)}$ are two disjoint subsets of vertices in G'_{n-2N-2} (possibly random but independent of the randomness in G'_{n-2N-2}). If $U^{(1)}$ and $U^{(2)}$ are sufficiently large, each of them is likely to contain at least one vertex from the giant component. Hence we may choose $N_1 \in \mathbb{N}$ such that, given the event $\{|U^{(1)}|, |U^{(2)}| \geq N_1\}$, there will exist a pair of vertices $u_1 \in U^{(1)}, u_2 \in U^{(2)}$ connected by a path in G'_{n-2N-2} of length at most $N_1 \log n$, with conditional probability at least $1 - \varepsilon$ for n sufficiently large.

Since the sizes $|\mathscr{U}^{(j)}|$ are independent Poisson random variables with mean εN, we may choose N large enough that $|\mathscr{U}^{(j)}| \geq N_1$ with probability at least $1 - \varepsilon$. Moreover, since $\mathcal{E}'_{2N}, \mathscr{U}^{(1)}, \mathscr{U}^{(2)}$ are finite and do not depend on n, it follows that $\mathcal{A}_n$ occurs with high probability and we can choose N_2 large enough that the diameters of $\mathcal{E}'_{2N}, \mathscr{U}^{(1)}, \mathscr{U}^{(2)}$ are at most N_2 with probability at least $1 - \varepsilon$.

Because of conditional independence, we may choose the vertex sets $U^{(j)} = \pi_M(\mathscr{U}^{(j)})$ in the preceding discussion. Taking the intersection of all the events above, it follows that, with probability at least $1 - 2\varepsilon - o(1)$, there is a path in K_n between vertices 1 and 2 consisting of at most N_2 edges of FPP weight at most $f_n(X_{\max}^{(1)})$; a single edge of weight at most $f_n(X_{\max}^{(1)} + \varepsilon)$; at most $N_1 \log n$ edges of weight at most $g(\frac{1}{n}(1 + \varepsilon)) = f_n(1 + \varepsilon)$; a single edge of weight at most $f_n(X_{\max}^{(2)} + \varepsilon)$; and at most N_2 edges of weight at most $f_n(X_{\max}^{(2)})$. Therefore, with probability at least $1 - 2\varepsilon - o(1)$,

$$\begin{aligned}
W_n &\leq (2N_2 + 2 + N_1 \log n) f_n\left((X_{\max}^{(1)} \vee X_{\max}^{(2)} \vee 1) + \varepsilon\right) \\
&\leq (2N_2 + 2 + N_1 \log n) f_n\left((M^{(1)} \vee M^{(2)}) + \varepsilon\right).
\end{aligned} \tag{4.1}$$

To complete the proof, it suffices to show that the right-hand side of (4.1) is at most $f_n((M^{(1)} \vee M^{(2)}) + 2\varepsilon)$ with high probability. Since $M^{(1)}$ and $M^{(2)}$ do not depend on n and satisfy $M^{(j)} \geq 1$ a.s., it suffices to show that, for each fixed $x \geq 1$, we have $(2N_2 + 2 + N_1 \log n) f_n(x) \leq f_n(x + \varepsilon)$ for n sufficiently large. But the assumptions on f_n

in (2.5) imply that the ratio $f_n(x+\varepsilon)/(2N_2+2+N_1\log n)f_n(x)$ diverges to infinity as $n\to\infty$ and therefore exceeds 1 for n sufficiently large. $\qquad\square$

5 Consequences of the scaling assumptions on f_n

In this section, we state the principal technical estimates that follow from Condition 2.2, as well as two results relating the function f_n to the assumptions on F_Y from Section 1.1.

Condition 2.2 gives a uniform lower bound on f_n'. If we rearrange and integrate this bound, we conclude that for any $m_0 > 1$, there exists $\delta > 0$ such that

$$\frac{f_n(x_1)}{f_n(x_0)} \geq \exp\left(\delta s_n \log(x_1/x_0)\right) = \left(\frac{x_1}{x_0}\right)^{\delta s_n} \qquad \text{for } \xi_n \leq x_0 \leq x_1 \leq m_0. \tag{5.1}$$

Lemma 5.1. *Suppose $(f_n(x))_n$ and $(s_n)_n$ satisfy Condition 2.2. Then, given $m_0 \in (1,\infty)$, there exists $K < \infty$ such that*

$$f_n^{-1}(y_1) - f_n^{-1}(y_0) \leq \frac{K}{s_n} \cdot \frac{y_1 - y_0}{y_0} \qquad \text{for } f_n(1) \leq y_0 \leq y_1 \leq f_n(m_0). \tag{5.2}$$

Lemma 5.2. *Suppose $(f_n(x))_n$ and $(s_n)_n$ satisfy Condition 2.2. Then, given $m_0 \in (1,\infty)$, there exists $K < \infty$ such that*

$$\left(f_n^{-1}(y_1) - f_n^{-1}(y_0)\right)^2 \leq \frac{K}{s_n^2} \cdot \frac{y_1 - y_0}{y_0} \qquad \text{for } f_n(1) \leq y_0 \leq y_1 \leq f_n(m_0). \tag{5.3}$$

Proof of Lemmas 5.1–5.2. Condition 2.2 implies that f_n is differentiable with positive derivative, and therefore continuous and strictly increasing, on $[1,\infty]$. In particular, we can replace y_i by $f_n(x_i)$, $i=0,1$, where $1 \leq x_0 \leq x_1 \leq m_0$. By (5.1),

$$\frac{f_n(x_1) - f_n(x_0)}{f_n(x_0)} \geq \exp\left(\delta s_n \log(x_1/x_0)\right) - 1 \geq \exp\left(\delta s_n (x_1-x_0)/m_0\right) - 1 \tag{5.4}$$

since $\frac{d}{dx}\log x \geq 1/m_0$ on $[1,m_0]$. Apply the inequalities $e^z - 1 \geq z$ and $e^z - 1 \geq \frac{1}{2}z^2$ and rearrange to prove both statements. $\qquad\square$

Lemma 5.3. *Suppose $(f_n(x))_n$ and $(s_n)_n$ satisfy Condition 2.2. Then, given any $m_0 \in [1,\infty)$, there exists $K \in (0,\infty)$ such that*

$$\int_0^m f_n(x)dx \leq \frac{Kf_n(m)}{s_n} \qquad \text{for all } m \in [1,m_0] \text{ and } n \in \mathbb{N}. \tag{5.5}$$

Proof. Apply (5.1) with $x_1 = m$ to conclude

$$\int_0^m f_n(x)dx \leq \int_0^{\xi_n} f_n(\xi_n)dx + \int_{\xi_n}^m f_n(m)\left(\frac{x}{m}\right)^{\delta s_n} dx$$

$$\leq f_n(\xi_n) + \frac{f_n(m)}{\delta s_n + 1} \leq f_n(\xi_n) + \delta^{-1}s_n^{-1}f_n(m). \tag{5.6}$$

Since $f_n(\xi_n) = o(f_n(1)/s_n)$, and *a fortiori* $f_n(\xi_n) = o(f_n(m)/s_n)$ for any $m \geq 1$, the result follows. $\qquad\square$

The next lemma shows that the hypotheses of Theorems 1.1–1.3 imply those of Theorems 2.1 and 2.11:

Lemma 5.4. *With the notation of (2.2)–(2.3), the relation (1.3) (with $t \mapsto L(t)$ slowly varying as $t \to \infty$) holds if and only if*

$$x\frac{d}{dx}\log g(x) = x^{-\alpha}\tilde{L}(1/x) \tag{5.7}$$

with $t \mapsto \tilde{L}(t)$ slowly varying as $t \to \infty$. If either of these two equivalent conditions hold, then $L(t) \sim \tilde{L}(t)$ as $t \to \infty$ and the sequences $(s_n)_n$, $(\tilde{s}_n)_n$ defined by (1.4) and

$$\tilde{s}_n = \frac{f'_n(1)}{f_n(1)} \tag{5.8}$$

satisfy $s_n \sim \tilde{s}_n$ as $n \to \infty$. Moreover if in addition $s_n \to \infty$ (or equivalently $\tilde{s}_n \to \infty$) then Conditions 2.2–2.3 hold for the sequences $(f_n(x), s_n)$ or $(f_n(x), \tilde{s}_n)$, and (2.15) holds. If the stronger statement $s_n/\log\log n \to \infty$ (or equivalently $\tilde{s}_n/\log\log n \to \infty$) holds, then (2.5) holds.

Proof. The equivalence follows by noting that $\widetilde{L}(1/x) = \mathrm{e}^{-x}(\frac{x}{1-\mathrm{e}^{-x}})^{1+\alpha}L(1/(1-\mathrm{e}^{-x}))$, so that $L(t) \sim \tilde{L}(t)$ as $t \to \infty$ and $\widetilde{L}(t)$ is slowly varying as $t \to \infty$ if and only if $L(t)$ is. As observed after Conditions 2.2–2.3, we have

$$\frac{xf'_n(x)}{f_n(x)} = \frac{(x/n)g'(x/n)}{g(x/n)} \tag{5.9}$$

so $\tilde{s}_n$ is obtained by setting $x = 1/n$ in (5.7). Since s_n is obtained by setting $u = 1/n$ in (1.3), $\tilde{s}_n \sim s_n$ follows from $L(t) \sim \tilde{L}(t)$.

We prove Conditions 2.2–2.3 for the sequence $(f_n(x), \tilde{s}_n)$. Since Conditions 2.2–2.3 are insensitive to replacing s_n by an asymptotically equivalent sequence, this will prove the result for the sequence $(f_n(x), s_n)$ also.

We compute

$$\frac{1}{\tilde{s}_n}\frac{xf'_n(x)}{f_n(x)} = \frac{g(1/n)}{(1/n)g'(1/n)}\frac{\frac{x}{n}g'(\frac{x}{n})}{g(\frac{x}{n})} = x^{-\alpha}\frac{\widetilde{L}(n/x)}{\widetilde{L}(n)}. \tag{5.10}$$

By properties of slowly-varying functions, the right-hand side of (5.10) is bounded away from 0 as $n \to \infty$, and indeed converges to $x^{-\alpha}$, uniformly over x in a compact subset of $(0, \infty)$. Set $\xi_n = 1/2$, say. To prove $\tilde{s}_n f_n(1/2) = o(f_n(1))$, we can apply (5.1) (which only uses the assumption that (5.10) is bounded away from 0) with $x_0 = 1/2, x_1 = 1$ to find

$$\frac{f_n(1)}{\tilde{s}_n f_n(1/2)} \geq \frac{2^{\delta\tilde{s}_n}}{\tilde{s}_n} \to \infty \tag{5.11}$$

since $\tilde{s}_n \to \infty$. This proves Condition 2.2.

To prove Condition 2.3, note that (5.10) is also bounded above uniformly over x in a compact subset of $(0, \infty)$. Since $\tilde{s}_n \to \infty$, we may choose η sufficiently small that the intervals $[1 - \eta/\tilde{s}_n, 1 + C/\tilde{s}_n]$ belong to such a compact subset for all n. Similarly to (5.1), we obtain the matching upper bound

$$\frac{f_n(1 + C/\tilde{s}_n)}{f_n(1 - \eta/\tilde{s}_n)} \leq \left(\frac{1 + C/\tilde{s}_n}{1 - \eta/\tilde{s}_n}\right)^{K\tilde{s}_n}. \tag{5.12}$$

Since $\tilde{s}_n \to \infty$, the right-hand side converges to $\mathrm{e}^{K(C+\eta)}$ as $n \to \infty$.

When $s_n, \tilde{s}_n \to \infty$, (2.15) follows immediately from (5.1). When $s_n/\log\log n \to \infty$, (2.5) follows from (5.1) because $\exp(\delta s_n \log((x + \delta)/x) - \log\log n) \to \infty$. $\quad\square$

The last result, valid without any scaling assumptions on f_n, states that scaling by f_n^{-1} or by nF_Y are equivalent.

Lemma 5.5. *With the notation of (2.2)–(2.3), where F_Y may depend implicitly on n, and with $(0, \infty)$-valued random variables W_n and $M^{(1)}$,*

$$f_n^{-1}(W_n) \overset{\text{P}}{\longrightarrow} M^{(1)} \quad \text{if and only if} \quad nF_Y(W_n) \overset{\text{P}}{\longrightarrow} M^{(1)}. \tag{5.13}$$

The same result is true if convergence in probability is replaced by convergence in distribution or convergence a.s.

Proof. It suffices to show that for every sequence (w_n) of finite positive numbers and every finite positive number m, $f_n^{-1}(w_n) \to m$ if and only if $nF_Y(w_n) \to m$. This is immediate from the relations

$$nF_Y(w_n) = n\left(1 - \mathrm{e}^{-f_n^{-1}(w_n)/n}\right), \qquad f_n^{-1}(w_n) = -n\log\left(1 - F_Y(w_n)\right) \tag{5.14}$$

(which follow from (2.2)–(2.3)) and the asymptotics $1 - \mathrm{e}^{-x} \sim x \sim -\log(1-x)$ as $x \downarrow 0$, together with the observation that, since m is finite, both $f_n^{-1}(w_n)/n$ and $F_Y(w_n)$ must tend to 0 as $n \to \infty$. $\qquad\square$

6 Poisson Galton–Watson trees and lucky vertices

In this section, we prove Proposition 2.17 and Lemma 2.18. We begin with preliminary results on Poisson Galton–Watson trees.

6.1 Properties of Poisson Galton–Watson trees

Proposition 6.1. *Write $\theta(m) = \mathbb{P}(|\mathrm{PGW}_m| = \infty)$ for the survival probability of a Poisson Galton–Watson tree of mean m. For $r \in (0, \infty)$, we denote by $\mathrm{PGW}_m^{(\leq r)}$ and $\mathrm{PGW}_m^{(\geq r)}$ the subgraph of PGW_m consisting of the vertices within distance r and with distance at least r, respectively, from the root.*

(a) *$\theta\colon (0, \infty) \to [0, 1]$ is non-decreasing with $\theta(m) = 0$ for $m \leq 1$, $\theta(m) > 0$ for $m > 1$, and*

$$1 - \theta(m) = \mathrm{e}^{-m\theta(m)} \qquad \text{for all } m \in (0, \infty). \tag{6.1}$$

(b) *As $m \downarrow 1$,*

$$\theta(m) \sim 2(m-1) \qquad \text{and} \qquad 1 - m(1 - \theta(m)) \sim m - 1, \tag{6.2}$$

and, uniformly over $m \geq 1$, $\theta(m) = O(m-1)$ and $1 - m(1 - \theta(m)) = O(m-1)$.

(c) *The derivative $\theta'(m)$ exists and satisfies $\theta'(m) \leq 2$ uniformly in $m > 1$.*

(d) *For $m > 1$, $\mathbb{P}\left(\mathrm{PGW}_m \in \cdot \,\middle|\, |\mathrm{PGW}_m| < \infty\right) = \mathbb{P}(\mathrm{PGW}_{\hat{m}} \in \cdot)$, where $\hat{m} = m(1 - \theta(m)) < 1$ and*

$$1 - \hat{m} \sim m - 1 \qquad \text{as } m \downarrow 1. \tag{6.3}$$

(e) *Uniformly over $m \in (0, \infty)$,*

$$\mathbb{P}(|\mathrm{PGW}_m| = k) = \frac{1}{m\sqrt{2\pi k^3}} \mathrm{e}^{-(m-1-\log m)k + O(1)}. \tag{6.4}$$

(f) *Let $K \in (0, \infty)$. Then there exists $c \in (0, \infty)$ such that for every $m \geq 1/K$ and every $k \in \mathbb{N}$ with $k \leq K/|m-1|^2$,*

$$\mathbb{P}(k \leq |\mathrm{PGW}_m| \leq 2k) \geq \frac{c}{\sqrt{k}}. \tag{6.5}$$

(g) Let $\varepsilon > 0$. Then there exists $\delta_1 > 0$ such that for all $r, k \in \mathbb{N}$ with $r^2 \geq \varepsilon k$, and for all $m > 0$,

$$\mathbb{P}(|\mathsf{PGW}_m^{(\leq r)}| \geq k) \geq \delta_1 \mathbb{P}(k \leq |\mathsf{PGW}_m| \leq 2k). \tag{6.6}$$

(h) Uniformly in $m, r \in [1, \infty)$,

$$\mathbb{P}\left(|\mathsf{PGW}_m^{(\geq r)}| \geq 1\right) \leq \theta(m) + O(1/r). \tag{6.7}$$

Proof. (a) Identity (6.1) is obtained by considering the individuals in the first generation. The remaining statements are standard facts about survival probabilities of Galton–Watson processes.

(b) For $\theta(m) > 0$, i.e., $m > 1$, we can solve (6.1) for m in terms of θ:

$$m = \frac{-\log(1 - \theta(m))}{\theta(m)}. \tag{6.8}$$

The function $\theta \mapsto \frac{-\log(1-\theta)}{\theta}$ increases from 1 to ∞ as θ increases from 0 to 1, and hence $\lim_{m \downarrow 1} \theta(m) = 0$. Expanding $-\log(1 - \theta)$ as a Taylor series around $\theta = 0$ gives $m - 1 = \frac{1}{2}\theta(m) + O(\theta(m)^2) \sim \frac{1}{2}\theta(m)$ and similarly $1 - m(1 - \theta(m)) = (\theta(m) + (1 - \theta(m))\log(1 - \theta(m)))/\theta(m) \sim \frac{1}{2}\theta(m)$. The uniform bounds follow because $\frac{\theta(m)}{m-1}$ and $\frac{1-m(1-\theta(m))}{m-1}$ are continuous and vanish in the limit $m \to \infty$.

(c) By (6.8), the function $m \mapsto \theta(m)$, $1 < m < \infty$, is the inverse of the function $\theta \mapsto \frac{-\log(1-\theta)}{\theta}$, $0 < \theta < 1$, so $\theta'(m) = 1/(\frac{dm}{d\theta})$. We compute

$$\frac{dm}{d\theta} = \frac{1}{\theta^2}\left(\frac{\theta}{1-\theta} + \log(1-\theta)\right) = \frac{1}{\theta^2}\int_0^\theta \left(\frac{1}{(1-u)^2} - \frac{1}{1-u}\right) du = \frac{1}{\theta^2}\int_0^\theta \frac{u}{(1-u)^2} du. \tag{6.9}$$

The last integral is bounded below by $\int_0^\theta u\, du = \frac{1}{2}\theta^2$ so $\frac{dm}{d\theta} \geq 1/2$ and $\theta'(m) \leq 2$.

(d) See for example Theorem 3.15 in [19]. The asymptotics in (6.3) follows from (6.2).

(e) It is well known (see for example the first display on page 951 in [4]) that

$$\mathbb{P}(|\mathsf{PGW}_m| = k) = \frac{e^{-mk}(mk)^{k-1}}{k!}. \tag{6.10}$$

Stirling's formula yields the claim.

(f) A Taylor expansion shows that $m - 1 - \log m \leq 2(m-1)^2$ for $1/2 \leq m \leq 3/2$, so that $(m - 1 - \log m)k \leq 2K$ for all the applicable values of k, m. Hence the probability in (6.4) is at least $c'/\sqrt{k^3}$ for some small $c' > 0$. Replacing k by i and summing over $i \in [k, 2k]$ yields the claim.

(g) Denote by $(\mathscr{G}_i, \mathbb{P})$ the uniform labelled rooted tree on i nodes after the labels of the children have been discarded and by $h(\mathscr{G}_i)$ the height of $\mathscr{G}_i$. Then $h(\mathscr{G}_i)/\sqrt{i}$ converges, as $i \to \infty$, to the maximum of $2B$ where $B = (B_t)_{t \in [0,1]}$ is a standard Brownian excursion [5, 6].

For all $m > 0$ and $i \in \mathbb{N}$, the distribution of PGW_m conditioned on having i nodes is the same as the distribution of $\mathscr{G}_i$. Letting $\mathscr{G}_i^{(\leq r)}$ denote the subtree of $\mathscr{G}_i$ consisting of vertices within distance r, we deduce that for all $i \leq 2k$,

$$\mathbb{P}(|\mathsf{PGW}_m^{(\leq r)}| = i) \geq \mathbb{P}(|\mathsf{PGW}_m| = i)\mathbb{P}(h(\mathscr{G}_i) \leq r)$$

$$\geq \mathbb{P}(|\mathsf{PGW}_m| = i)\mathbb{P}\left(\frac{h(\mathscr{G}_i)}{\sqrt{i}} \leq \frac{r}{\sqrt{i}}\right)$$

$$\geq \mathbb{P}(|\mathsf{PGW}_m| = i)\mathbb{P}\left(\frac{h(\mathscr{G}_i)}{\sqrt{i}} \leq \frac{r}{\sqrt{2n}}\right)$$

$$\geq \mathbb{P}(|\mathsf{PGW}_m| = i)\mathbb{P}\left(\frac{h(\mathscr{G}_i)}{\sqrt{i}} \leq \sqrt{\varepsilon/2}\right). \tag{6.11}$$

Write p_i for the last probability on the right-hand side of (6.11). Then $\lim_{i \to \infty} p_i > 0$ and there is some $\delta > 0$ such that $p_i \geq \delta$ for all sufficiently large i. Since $\mathbb{P}(h(\mathscr{G}_i) \leq r) > 0$ for each i, we can assume that $\mathbb{P}(|\mathsf{PGW}_m^{(\leq r)}| = i) \geq \delta_1 \mathbb{P}(|\mathsf{PGW}_m| = i)$ for all i by taking δ_1 sufficiently small. Summing over $i \in [k, 2k]$ yields the claim.

(h) Using (d), we obtain

$$
\begin{aligned}
\mathbb{P} &\left(|\mathsf{PGW}_m^{(\geq r)}| \geq 1 \right) \\
&= \mathbb{P}(|\mathsf{PGW}_m| = \infty) + \mathbb{P}(|\mathsf{PGW}_m| < \infty) \mathbb{P}\left(|\mathsf{PGW}_m^{(\geq r)}| \geq 1 \,\big|\, |\mathsf{PGW}_m| < \infty \right) \\
&= \theta(m) + (1 - \theta(m)) \mathbb{P}(|\mathsf{PGW}_{\hat{m}}^{(\geq r)}| \geq 1) \\
&\leq \theta(m) + \mathbb{P}(|\mathsf{PGW}_1^{(\geq r)}| \geq 1).
\end{aligned}
\tag{6.12}
$$

The claim now follows from a standard result on critical Galton–Watson processes (see for example [18, Lemma I.10.1]). $\qquad\square$

6.2 The probability of luckiness: proof of Proposition 2.17

In the proof of Proposition 2.17 we use the following estimates:

Lemma 6.2. *Suppose $0 \leq X \leq Y$ are random variables such that $\mathbb{E}(X \,|\, Y) \geq pY$ a.s. Then $\mathbb{P}(X \geq m) \geq \frac{1}{2}p \mathbb{P}(Y \geq 2m/p)$ for any $m \in [0, \infty)$.*

Proof. Markov's inequality applied to $Y - X$ gives $\mathbb{P}\left(Y - X > (1 - \frac{1}{2}p)Y \,\big|\, Y \right) \leq (1 - p)/(1 - \frac{1}{2}p)$, so $\mathbb{P}\left(X \geq \frac{1}{2}pY \,\big|\, Y \right) \geq \frac{1}{2}p/(1 - \frac{1}{2}p) \geq \frac{1}{2}p$ and the result follows. $\qquad\square$

For later use, it will be convenient to prove a slight strengthening of Proposition 2.17:

Lemma 6.3. *Suppose $(f_n(x))_n$ and $(s_n)_n$ satisfy Condition 2.2, and let $m_0 \in (1, \infty)$ and $K \in (0, \infty)$. Then there exists $\delta > 0$ such that for all $m \in [1, m_0]$ and for all $k, n \in \mathbb{N}$ with $k \leq K(s_n^2 \wedge |m - 1|^{-2})$,*

$$
\mathbb{P}\left(|\mathsf{BP}_{f_n(1)}| \geq k \,\big|\, |\mathsf{PGW}_m| < \infty \right) \geq \frac{\delta}{\sqrt{k}}.
\tag{6.13}
$$

When $m = 1$, we interpret $|m - 1|^{-2}$ as ∞, so that the condition on k reduces to $k \leq K s_n^2$.

Proof. Let $m_0 \in (1, \infty)$ and $K \in (0, \infty)$ be given and let $m \in [1, m_0]$. We consider PGW_m as a rooted labelled tree equipped with edge weights, but with the vertex labels from the PWIT forgotten.[3] Then, conditional on PGW_m, the PWIT edge weights X_w, $w \in \mathsf{PGW}_m \setminus \{\varnothing_1\}$, are uniformly distributed on $[0, m]$, conditionally independent across different edge weights.

By Proposition 6.1 (d), the conditional distribution of PGW_m given $\{|\mathsf{PGW}_m| < \infty\}$ is that of a Poisson Galton–Watson tree with mean $\hat{m}$. Under this conditioning, the PWIT edge weights along edges of PGW_m are still uniformly distributed on $[0, m]$. So the conditional law of PGW_1 given $\{|\mathsf{PGW}_m| < \infty\}$ is that of a Poisson Galton–Watson tree with mean

$$
m' = \frac{\hat{m}}{m}.
\tag{6.14}
$$

From Proposition 6.1 (d),

$$
1 - m' \sim 2(m - 1) \qquad \text{as } m \downarrow 1,
\tag{6.15}
$$

---

[3]Formally, we should consider instead of PGW_m the subset $\mathsf{P\tilde{G}W}_m$ where we replace each vertex $w \in \mathsf{PGW}_m \setminus \{v\}$ by an arbitrary label $\ell(w)$ drawn independently from some continuous distribution. By a slight abuse of notation, we will refer to PGW_m and X_w, $w \in \mathsf{PGW}_m \setminus \{v\}$ instead of $\mathsf{P\tilde{G}W}_m$ and $X_{\ell^{-1}(w)}$, $w \in \mathsf{P\tilde{G}W}_m$. This procedure avoids the complication, implicit in our Ulam–Harris notation, that the vertex $w = \varnothing_1 k_1 k_2 \ldots k_r \in \mathcal{T}$ automatically gives information about the number of its siblings with smaller edge weights.

and therefore there is a constant K' such that $k \leq K\,|m-1|^{-2}$ implies $k \leq K'\,|1-m'|^{-2}$. Increasing K' if necessary, we may also assume that $m' \geq 1/K'$ for $m \leq m_0$.

Let C be the constant from Lemma 5.3 with $m = 1$, and define

$$r = \frac{s_n}{2C}. \tag{6.16}$$

By Proposition 6.1 (f) and (g) (with m replaced by m' and K replaced by K', and with $\varepsilon = 1/(16C^2K)$, respectively) there are constants $c, \delta_1 > 0$ such that

$$\mathbb{P}\left(|\mathsf{PGW}_1^{(\leq r)}| \geq 4k \,\big|\, |\mathsf{PGW}_m| < \infty\right) \geq \delta_1 \mathbb{P}(4k \leq |\mathsf{PGW}_{m'}| \leq 8k) \geq \frac{\delta_1 c}{\sqrt{4k}} \tag{6.17}$$

whenever $r^2 \geq 4\varepsilon k$, or equivalently (by the definition of ε and r) whenever $k \leq Ks_n^2$.

Conditional on PGW_1, the PWIT edge weights X_w, $w \in \mathsf{PGW}_1 \setminus \{\varnothing_1\}$, are uniformly distributed on $[0, 1]$. Therefore the first passage edge weights $Y_w = f_n(X_w)$ satisfy

$$\mathbb{E}\left(Y_w \,\big|\, w \in \mathsf{PGW}_1 \setminus \{\varnothing_1\}, |\mathsf{PGW}_m| < \infty\right) = \int_0^1 f_n(x)dx \leq \frac{Cf_n(1)}{s_n}. \tag{6.18}$$

By the definition of $\mathsf{PGW}_1^{(\leq r)}$, the height of $v \in \mathsf{PGW}_1^{(\leq r)}$ is at most r. Hence, (6.18) and (6.16) give

$$\mathbb{E}\left(T_v \,\big|\, v \in \mathsf{PGW}_1^{(\leq r)}, |\mathsf{PGW}_m| < \infty\right) \leq r\frac{Cf_n(1)}{s_n} \leq \tfrac{1}{2}f_n(1). \tag{6.19}$$

By Markov's inequality $\mathbb{P}\left(T_v > f_n(1) \,\big|\, v \in \mathsf{PGW}_1^{(\leq r)}, |\mathsf{PGW}_m| < \infty\right) \leq \tfrac{1}{2}$ and consequently

$$\mathbb{P}\left(v \in \mathsf{BP}_{f_n(1)} \,\big|\, v \in \mathsf{PGW}_1^{(\leq r)}, |\mathsf{PGW}_m| < \infty\right) \geq \frac{1}{2}, \tag{6.20}$$

so that

$$\mathbb{E}\left(\left|\mathsf{BP}_{f_n(1)}\right| \,\big|\, \left|\mathsf{PGW}_1^{(\leq r)}\right|\right) \geq \frac{1}{2}\left|\mathsf{PGW}_1^{(\leq r)}\right| \qquad \text{on } \{|\mathsf{PGW}_m| < \infty\}. \tag{6.21}$$

By Lemma 6.2 with $X = \left|\mathsf{BP}_{f_n(1)}\right|$, $Y = \left|\mathsf{PGW}_1^{(\leq r)}\right|$ and $p = 1/2$, we obtain

$$\mathbb{P}\left(\left|\mathsf{BP}_{f_n(1)}\right| \geq k \,\big|\, |\mathsf{PGW}_m| < \infty\right) \geq \frac{1}{4}\mathbb{P}\left(\left|\mathsf{PGW}_1^{(\leq r)}\right| \geq 4k \,\big|\, |\mathsf{PGW}_m| < \infty\right)$$

$$\geq \frac{\delta_1 c}{4\sqrt{4k}}. \qquad \square$$

Proof of Proposition 2.17. Clearly $\mathbb{P}(v \text{ is } k\text{-lucky})$ does not depend on v, so Proposition 2.17 is the special case $m = 1$ in Lemma 6.3, where $|\mathsf{PGW}_m| < \infty$ already holds a.s. and the conditional probability reduces to an ordinary probability. $\qquad \square$

6.3 Emergence of lucky vertices

Define

$$T_{\text{late } k-\text{lucky}} = \inf\left\{T_v \colon v \in \mathcal{T} \setminus \{\varnothing_1\} \text{ is } k\text{-lucky and } T_v > T_{p(v)} + f_n(1)\right\}, \tag{6.22}$$

the first time that a lucky vertex is born to a parent of age greater than $f_n(1)$.

Lemma 6.4. *The distribution of*

$$\sum_{v \in \mathsf{BP}_{T_{\text{late } k-\text{lucky}}}} \left(f_n^{-1}\left(T_{\text{late } k-\text{lucky}} - T_v\right) - 1\right)^+ \tag{6.23}$$

is exponential with rate $\mathbb{P}(v \text{ is } k\text{-lucky})$.

Since $T_v - T_{p(v)} = f_n(X_v)$, the condition $T_v > T_{p(v)} + f_n(1)$ in the definition of $T_{\text{late } k-\text{lucky}}$ is equivalent to $X_v > 1$. On the other hand, the event $\{v \text{ is } k\text{-lucky}\}$ depends only on the evolution of $\text{BP}^{(v)}$ until time $f_n(1)$ and is therefore determined by those descendants v' of v for which $X_{v'} \leq 1$. Because these two conditions on edge weights are mutually exclusive, it will follow that $T_{\text{late } k-\text{lucky}}$ is the first arrival time of a certain Cox process. We now formalize this intuition, which requires some care.

Proof of Lemma 6.4. To avoid complications arising from our Ulam–Harris notation, we modify our description of vertices as follows. Instead of the vertex $v = \varnothing_1 k_1 k_2 \ldots k_r$ we will consider the modified vertex $w(v) = \varnothing_1 X_{k_1} X_{k_1 k_2} \ldots X_{k_1 k_2 \ldots k_r}$ formed out of the edge weights along the path from $\varnothing_1$ to v. We can extend our usual notation for parents, length, concatenation, edge weight, and birth times to vertices of the form $w = \varnothing_1 x_1 \cdots x_r$, $x_i \in (0, \infty)$: for instance, $|w| = r$, $X_w = x_r$ and $T_w = f_n(x_1) + \cdots + f_n(x_r)$.

Form the point measure $\mathcal{M} = \sum_{v \in \mathcal{T}} \delta_{w(v)}$ on $\cup_{r=0}^{\infty} \{\varnothing_1\} \times (0, \infty)^r$. Given $\mathcal{M}$, we can recover the PWIT $(\mathcal{T}, X)$: for instance, $X_{\varnothing_1 k_1} = \inf \{x > 0 \colon \mathcal{M}(\{\varnothing_1\} \times (0, x]) \geq k_1\}$ and $X_{\varnothing_1 k_1 k_2} = \inf \{x > 0 \colon \mathcal{M}(\{\varnothing_1\} \times \{X_{\varnothing_1 k_1}\} \times (0, x]) \geq k_2\}$. The point measure $\mathcal{M}$ has the advantage that a value such as $\mathcal{M}(\{\varnothing_1\} \times (a, b))$ (the number of children of $\varnothing_1$ with edge weights in the interval (a, b)) does not reveal information about the number of sibling edges of smaller edge weight.

The Poisson property of the PWIT can be expressed in terms of $\mathcal{M}$ by saying that, conditional on the restriction $\mathcal{M}|_{\{\varnothing_1\} \times (0, \infty)^r}$ to the first r generations, the $(r + 1)^{\text{st}}$ generation $\mathcal{M}|_{\{\varnothing_1\} \times (0, \infty)^{r+1}}$ is formed as a Cox process with intensity $\mathcal{M}|_{\{\varnothing_1\} \times (0, \infty)^r} \otimes \mathbb{1}_{\{x>0\}} dx$, where $\mathbb{1}_{\{x>0\}} dx$ denotes Lebesgue measure on $(0, \infty)$.

We next rearrange the information contained in $\mathcal{M}$. Given $w = \varnothing_1 x_1 \ldots x_r$, let $\mathcal{M}^{(w)} = \sum_{w' \colon \mathcal{M}(\{ww'\})=1} \delta_{w'}$ denote the point measure corresponding to all descendants of w (thus $\mathcal{M}^{(w)} = 0$ if $\mathcal{M}(\{w\}) = 0$, and the point measures $\mathcal{M}^{(w(v))}$, $v \in \mathcal{T}$, are identically distributed and non-zero)[4]. Further, write $\mathcal{M}^{(w)}_{\leq}$ for the restriction of $\mathcal{M}^{(w)}$ to $\cup_{k=0}^{\infty} (0, 1]^k$. Since luckiness only depends on descendants explored within time $f_n(1)$, it follows that whether or not v is k-lucky can be determined solely in terms of $\mathcal{M}^{(w(v))}_{\leq}$. Indeed, call a point measure $\mathfrak{m}$ on $\cup_{k=0}^{\infty} (0, 1]^k$ k-*lucky* if $\mathfrak{m}(\{w' \colon T_{w'} \leq f_n(1)\}) \geq s_n^2/\varepsilon_1$; then v is k-lucky if and only if $\mathcal{M}^{(w(v))}_{\leq}$ is k-lucky.

Define A to be the collection of vertices $w(v)$, $v \in \mathcal{T}$, that are born before time $T_{\text{late } k-\text{lucky}}$. That is, for any non-root ancestor v' of v, including v itself, it is *not* the case that $T_{v'} > T_{p(v')} + f_n(1)$ and v' is k-lucky. To study A, we decompose the PWIT according to vertices v' that are born late (i.e., $T_{v'} > T_{p(v')} + f_n(1)$) and keep track of their early explored descendants (i.e., $\mathcal{M}^{(w(v'))}_{\leq}$).

For $w = \varnothing_1 x_1 \ldots x_r$, let $i_1 < \cdots < i_k$ denote those indices (if any) for which $x_i > 1$, and write $w_\ell = \varnothing_1 x_1 \ldots x_{i_\ell}$, $\ell = 1, \ldots, k$. Set $q(w, \mathcal{M})$ to be the sequence of measures $\mathcal{M}^{(\varnothing_1)}_{\leq} \mathcal{M}^{(w_1)}_{\leq} \ldots \mathcal{M}^{(w_k)}_{\leq}$. Define the rearranged point measure

$$\mathcal{R} = \int_{\{\varnothing_1\} \cup \bigcup_{k=1}^{\infty} \{\varnothing_1\} \times (0, \infty)^{k-1} \times (1, \infty)} \delta_{(w, q(w, \mathcal{M}))} d\mathcal{M}(w). \qquad (6.24)$$

(That is, $\mathcal{R}$ is a point measure on pairs (w, q) such that w satisfies $w = \varnothing_1$ or $X_w > 1$, and q is a sequence of measures on $\cup_{r=0}^{\infty} (0, 1]^r$. Considering $\mathcal{R}$ instead of $\mathcal{M}$ corresponds to partitioning vertices according to their most recent ancestor (if any) having edge weight greater than 1.)

The Poisson property of the PWIT implies that, conditional on $\mathcal{R}|_{\{(w, q) \colon |w| \leq r\}}$, the

---

[4] In this instance, we assume that w' does not have a $\varnothing_j$ symbol for convenience of notation.

restriction $\mathcal{R}\big|_{\{(w,q)\colon |w|=r+1\}}$ forms a Cox process with intensity measure

$$
\int_{\{(w,q)\colon |w|\leq r\}} d\mathcal{R}(w,q) \int_{\{w'\colon |ww'|=r\}} d\mathcal{M}^{(w)}_{\leq}(w') \left[\delta_{ww'} \otimes \mathbb{1}_{\{x>1\}} dx\right] \otimes \left[\delta_q \otimes d\mathbb{P}(\mathcal{M}^{(\varnothing 1)}_{\leq} \in \cdot)\right],
$$
(6.25)

where $\delta_{ww'} \otimes \mathbb{1}_{\{x>1\}} dx$ means the image of Lebesgue measure on $(1,\infty)$ under the concatenation mapping $x \mapsto ww'x$. The formula (6.25) expresses the fact that every vertex in the $(r+1)^{\text{st}}$ generation has a parent uniquely written as ww', with $(w, q(w, \mathcal{M}))$ corresponding to a point mass in $\mathcal{R}$, w' corresponding to a point mass in the last entry $\mathcal{M}^{(w)}_{\leq}$ of the sequence $q(w, \mathcal{M})$, and $r = |w| + |w'|$.

Now, it is easy to verify that A is measurable with respect to the restriction of $\mathcal{R}$ to pairs (w,q) such that $q = \mathfrak{m}_0 \ldots \mathfrak{m}_k$ with $\mathfrak{m}_\ell$ not k-lucky for each $\ell \neq 0$.

Because of the Poisson property of the PWIT, as expressed via $\mathcal{R}$ in (6.25), it follows that, conditional on A, the point measure $L = \sum_{w'' \notin A, p(w'') \in A} \delta_{T_{w''}}$ forms a Cox process on $(0,\infty)$. Furthermore, the first point of L is precisely $T_{\text{late } k-\text{lucky}}$. To determine the intensity measure of L, we note that for a vertex $ww' \in A$, $w'' = ww'x$ satisfies $w'' \notin A$ if and only if $X_{w''} > 1$ and $\mathcal{M}^{(w'')}_{\leq}$ is k-lucky. Furthermore, the condition $T_{w''} = T_{ww'} + f_n(X_{w''}) \leq t$ is equivalent to $X_{w''} \leq f_n^{-1}(t - T_{ww'})$. Using (6.25), it follows that the cumulative intensity measure of L on $(0,t]$ is given by

$$
\int_{\left\{\substack{(w,q)\colon q=\mathfrak{m}_0\ldots\mathfrak{m}_k, \\ \mathfrak{m}_\ell \text{ not } k\text{-lucky for any } \ell\neq 0}\right\}} d\mathcal{R}(w,q) \int_{\{w'\colon T_{ww'}\leq t\}} d\mathcal{M}^{(w)}_{\leq}(w') \left(f_n^{-1}(t - T_{ww'}) - 1\right)^+ \quad (6.26)
$$
$$
\times \, \mathbb{P}(\mathcal{M}^{(\varnothing 1)}_{\leq} \text{ is } k\text{-lucky}).
$$

The vertices ww' from the integral in (6.26) are in one-to-one correspondence with the vertices $u \in A$ satisfying $T_u \leq t$. Consequently we may re-write the cumulative intensity as

$$
\sum_{u \in A\colon T_u \leq t} \left(f_n^{-1}(t - T_u) - 1\right)^+ \cdot \mathbb{P}(u \text{ is } k\text{-lucky}). \tag{6.27}
$$

Finally we note that $\mathbb{P}(v \text{ is } k\text{-lucky})$ times the sum in (6.23) is exactly the sum in (6.27) evaluated at $t = T_{\text{late } k-\text{lucky}}$. (The vertex v for which $T_v = T_{\text{late } k-\text{lucky}}$ does not contribute to (6.23).) The cumulative intensity in (6.27) is a.s. continuous as a function of t (since f_n^{-1} is continuous and the jumps at the times T_u are zero). But for any Cox process with continuous cumulative intensity function and infinite total intensity, it is elementary to verify that when the cumulative intensity is evaluated at the first point of the Cox process, the result is exponential with mean 1. This completes the proof. $\qquad \square$

Proof of Lemma 2.18. The argument in the proof of Lemma 6.4 shows that, conditionally on V and $(\mathrm{BP}_t)_{t\leq T_V+f_n(1)}$, the children of a vertex $v \in D$ born after time $T_v + f_n(1)$ appear as a Poisson point process, conditionally independent over the choice of $v \in D$. In particular, the number of children of a vertex $v \in D$ born in the time interval $(T_v + f_n(1), T_v + f_n(1 + C/s_n))$ is Poisson with mean C/s_n, conditionally independently over the choice of $v \in D$, and the number of those children that are Rs_n^2-lucky is Poisson with mean $C\mathbb{P}(w \text{ is } Rs_n^2\text{-lucky})/s_n$. Summing over $v \in D$, it follows that by time $T_V + f_n(1) + f_n(1 + C/s_n)$, the number of Rs_n^2-lucky children w born to vertices $v \in D$ after time $T_v + f_n(1)$ is at least as large as a Poisson random variable with mean

$$
\frac{C}{s_n}\mathbb{P}(w \text{ is } Rs_n^2\text{-lucky})(rs_n^2) = Crs_n\mathbb{P}(w \text{ is } Rs_n^2\text{-lucky}). \tag{6.28}
$$

By Proposition 2.17, the mean in (6.28) is at least δC, uniformly in n. Hence, given $\varepsilon > 0$ and $k \in \mathbb{N}$, we may ensure that a Poisson random variable with mean (6.28) has value at

least k with probability at least $1 - \varepsilon$ by taking $C > 0$ sufficiently large. We conclude that $T_{V_k} - T_V \leq f_n(1) + f_n(1 + C/s_n)$ with probability at least $1 - \varepsilon$. But this completes the proof because Condition 2.3 implies that $f_n(1 + C/s_n) = O(f_n(1))$ for each fixed C. $\qquad \square$

7 IP and the geometry of the exploration process

In this section, we compare the branching process $\mathrm{BP}_{\mathrm{BB}\,\sigma_n}$ to the IP cluster $\mathrm{IP}(\infty)$ – the set of all vertices ever invaded in the IP process. The structure of the IP cluster is encoded in a single infinite backbone and an associated process of maximum weights, with off-backbone branches expressed in terms of Poisson Galton–Watson trees. See Proposition 7.1 below.

The proofs in this section rely on showing that $\mathrm{BP}_{\mathrm{BB}\,\sigma_n}$ cannot be significantly larger than the part of $\mathrm{IP}(\infty)$ within distance σ_n along the backbone. Specifically, we show that (a) the time to explore a section of backbone is not much larger than the time to explore the largest edge along it (Lemma 7.3); (b) the number of uninvaded edges on the boundary of $\mathrm{IP}(\infty)$ that are explored by time $T_{\mathrm{BB}\,\sigma_n}$ is small (Lemma 7.6); (c) by time $T_{\mathrm{BB}\,\sigma_n}$, it is likely that $\delta\sigma_n^2$-lucky vertices have been explored, for $\delta > 0$ small (Lemma 7.9); and (d) the likelihood of exploring very long paths that do not belong to the IP cluster is moderate. Assertions (a)–(c) will allow us to prove Theorem 2.16, and assertion (d) will be made precise in Lemma 7.11 and in the proofs of Lemma 7.12 and Theorem 2.15.

7.1 Structure and scaling of the IP cluster

Our description of $\mathrm{IP}(\infty)$ is based on [4], which examines the structure of the IP cluster on the PWIT, and the scaling limit results in [8], which proves similar results for regular trees. As remarked in [4], the scaling limit results of [8] can be transferred to the PWIT without difficulty.

Recall the backbone and the backbone edge weights X_k^{BB} introduced in Definition 2.13. We define the forward maximum MC_k by

$$\mathsf{MC}_k = \sup_{i > k} X_i^{\mathrm{BB}}. \tag{7.1}$$

In the notation of Definition 2.13, the maximum invaded edge weight $M^{(1)}$ from (2.19) is now MC_0. (This amounts to the observation, elementary to verify, that the largest edge weight $M^{(1)}$ must occur as one of the backbone edge weights X_k^{BB}.)

The off-backbone branch at height k means the subtree of $\mathrm{IP}(\infty)$ consisting of vertices that are descendants of V_k^{BB} but not descendants of V_{k+1}^{BB}, and is denoted by τ_k. We consider τ_k as a rooted labelled tree, but with the edge weights and vertex labels from the PWIT forgotten[5].

Proposition 7.1 ([4, 8]). *The backbone is well defined, and $\mathsf{MC}_k > 1$ for each k, a.s. Furthermore:*

(a) *The maximum in (7.1) is attained uniquely, for each k, a.s. Writing I_k for the random height at which the maximum in (7.1) is attained, it holds that $I_k = O_{\mathbb{P}}(k \vee 1)$ uniformly over $k \in \mathbb{N}_0$.*

---

[5]As in the proof of Proposition 2.17, we should consider instead of τ_k the set $\tilde{\tau}_k$ where we replace each vertex $v \in \tau_k \setminus \{V_k^{\mathrm{BB}}\}$ by an arbitrary label $\ell(v)$ drawn independently from some continuous distribution. By a slight abuse of notation, we will refer to τ and X_v, $v \in \tau_k \setminus \{V_k^{\mathrm{BB}}\}$ instead of $\tilde{\tau}_k$ and $X_{\ell^{-1}(v)}$, $v \in \tilde{\tau}_k$. This procedure avoids the complication, implicit in our Ulam–Harris notation, that the vertex $v = \varnothing_1 k_1 k_2 \ldots k_r \in \mathcal{T}$ automatically gives information about the number of its siblings with smaller edge weights.

(b) *The sequence* $(\mathsf{MC}_k)_{k=0}^\infty$ *is non-increasing and forms a Markov chain with initial distribution* $\mathbb{P}(\mathsf{MC}_0 \le m) = \theta(m)$ *and transition mechanism*

$$\mathbb{P}\left(\mathsf{MC}_{k+1} = m \mid \mathsf{MC}_k = m\right) = m(1-\theta(m)), \qquad m > 1,$$

$$\mathbb{P}\left(\mathsf{MC}_{k+1} < m' \mid \mathsf{MC}_k = m\right) = \frac{\theta(m')}{\theta(m)}(1 - m(1-\theta(m))), \quad 1 \le m' \le m. \tag{7.2}$$

(c) $\mathsf{MC}_k = 1 + \Theta_\mathbb{P}(1/k)$ *and indeed* $k(\mathsf{MC}_k - 1)$ *converges weakly to an exponential distribution with mean* 1 *as* $k \to \infty$.

(d) *Conditionally on* $(\mathsf{MC}_k)_{k=0}^\infty$, *the off-backbone branches* $(\tau_k)_{k=0}^\infty$ *are distributed as subcritical Poisson Galton–Watson trees with means* $\mathsf{MC}_k(1-\theta(\mathsf{MC}_k))$, *conditionally independent (but not identically distributed) for each* k.

(e) *Conditionally on* $(\mathsf{MC}_k)_{k=0}^\infty$, *the PWIT edge weight* X_k^{BB} *either (i) equals* MC_{k-1}, *if* $\mathsf{MC}_k < \mathsf{MC}_{k-1}$; *or (ii) has the* $\text{Uniform}[0, \mathsf{MC}_k]$ *distribution, if* $\mathsf{MC}_k = \mathsf{MC}_{k-1}$. *Furthermore the weights are conditionally independent for each* k.

(f) *Conditionally on* $(\mathsf{MC}_k)_{k=0}^\infty$ *and* $(\tau_k)_{k=0}^\infty$, *the PWIT edge weight of an edge between two vertices of* τ_k *has the* $\text{Uniform}[0, \mathsf{MC}_k]$ *distribution, conditionally independent over the choice of edge and of* k.

(g) *Conditionally on* $(\mathsf{MC}_k)_{k=0}^\infty$ *and* $(\tau_k)_{k=0}^\infty$, *the collection of PWIT edge weights between a vertex* $v \in \tau_k$ *and all child vertices* vi *for which* $vi \notin \tau_k$, *forms a Poisson point process of intensity* 1 *on the interval* (MC_k, ∞). *Moreover these Poisson point processes are conditionally independent for every* k *and every* $v \in \tau_k$.

(h) *Conditionally on the IP cluster* $\text{IP}(\infty)$ *and all its internal and boundary edge weights, the edge weights* $(X_{wk})_{k\in\mathbb{N}}$ *associated to a vertex* $w \notin \text{IP}(\infty) \cup \partial\text{IP}(\infty)$ *form a Poisson point process of rate* 1, *conditionally independent over the choice of* w.

(i) *The part of* $\text{IP}(\infty)$ *not descended from* V_k^{BB} *has size* $O_\mathbb{P}(k^2)$ *and diameter* $O_\mathbb{P}(k)$.

Proof. The backbone is well defined by Corollary 22 in [4]. The same paper proves (a) in Theorems 21 and 30, (b) in Section 3.3, (d) in Theorem 31, and (e) in Theorem 3. It has been observed on the top of page 954 in [4] that the methodology of [8] can be applied to show that [8, Proposition 3.3] holds for the PWIT, proving (c).

For parts (f) and (g), notice that the event that τ_k equals a particular finite tree τ requires that the children of V_k^{BB} should consist of (i) the child V_{k+1}^{BB} with edge weight consistent with the process $(\mathsf{MC}_l)_{l\ge0}$; and (ii) other children, and their descendants, joined to V_k^{BB} by edges of weight less than MC_k, in numbers corresponding to the structure specified by τ. However, conditioning on $(\mathsf{MC}_k)_{k=0}^\infty$ and $(\tau_k)_{k=0}^\infty$ does not impose any constraint on the precise value of the edge weights less than MC_k, the uninvaded edge weights that exceed MC_k, or the edge weights outside of $\text{IP}(\infty)$ and its boundary. Parts (f), (g) and (h) therefore follow from properties of Poisson point processes.

For the bound on the size in (i), it suffices to note that the expected size conditionally on $(\mathsf{MC}_{k'})_{k'=0}^\infty$ is at most $k \cdot 1/(1 - \widehat{\mathsf{MC}}_k)$ by (d) and the formula for the total expected offspring in a Galton–Watson tree of mean $\widehat{\mathsf{MC}}_k$. By (c), the conditional expectation of this size is $O_\mathbb{P}(k^2)$.

For the bound on the diameter, it suffices to notice that (d) implies that the maximum distance from the root to a vertex of $\bigcup_{j=0}^{k-1} \tau_j$ is stochastically dominated by k plus the maximum of k extinction times from k independent critical Poisson Galton–Watson branching

processes. Since the probability that a critical Poisson Galton–Watson branching process lives to generation ℓ is $O(1/\ell)$, the claim follows. $\qquad\square$

We next give a lemma for later use in bounding expectations of functions of the backbone edge weights X_k^{BB}.

Lemma 7.2. *There is a constant $C < \infty$ such that, for every non-negative measurable function h and every $k, k_0 \in \mathbb{N}$ with $k_0 < k$,*

$$\mathbb{E}\left(h(X_k^{\mathsf{BB}})\mathbb{1}_{\left\{X_k^{\mathsf{BB}} < \mathsf{MC}_{k_0}\right\}}\,\Big|\,\mathsf{MC}_{k_0} = m\right) \le C \int_0^m h(x)dx. \tag{7.3}$$

Proof. There are three possibilities: either (a) $\mathsf{MC}_k < \mathsf{MC}_{k-1} = \mathsf{MC}_{k_0}$; (b) $\mathsf{MC}_k = \mathsf{MC}_{k-1}$; or (c) $\mathsf{MC}_k < \mathsf{MC}_{k-1} < \mathsf{MC}_{k_0}$. Recalling Proposition 7.1 (e), we see that in case (a) we have $X_k^{\mathsf{BB}} = \mathsf{MC}_{k-1} = \mathsf{MC}_{k_0}$, which does not contribute to the expectation. In case (b), the conditional distribution of X_k^{BB} is $\mathrm{Uniform}[0, \mathsf{MC}_k]$ and

$$\mathbb{E}\left(h(X_k^{\mathsf{BB}})\,\big|\,\mathsf{MC}_k = \mathsf{MC}_{k-1}, \mathsf{MC}_{k_0} = m\right)$$
$$= \mathbb{E}\left(\int_0^{\mathsf{MC}_k} h(x)dx/\mathsf{MC}_k\,\bigg|\,\mathsf{MC}_k = \mathsf{MC}_{k-1}, \mathsf{MC}_{k_0} = m\right) \le \int_0^m h(x)dx \tag{7.4}$$

since $1 \le \mathsf{MC}_k \le m$. For case (c), define $K' = \max\left\{k'\colon \mathsf{MC}_{k'} > \mathsf{MC}_{k-1}\right\}$, which is well defined since $\mathsf{MC}_{k_0} > \mathsf{MC}_{k-1}$. Partition according to the value of K', noting that

$$\{K' = k'\} = \left\{\mathsf{MC}_{k-1} = \mathsf{MC}_{k'+1} < \mathsf{MC}_{k'}\right\}, \qquad \{\mathsf{MC}_{k-1} < \mathsf{MC}_{k_0}\} = \left\{k_0 \le K' < k-1\right\}. \tag{7.5}$$

Since $X_k^{\mathsf{BB}} = \mathsf{MC}_{k-1}$ under case (c), we find

$$\mathbb{E}\left(h(X_k^{\mathsf{BB}})\mathbb{1}_{\left\{\mathsf{MC}_k < \mathsf{MC}_{k-1} < \mathsf{MC}_{k_0}\right\}}\,\Big|\,\mathsf{MC}_{k_0} = m\right)$$
$$= \sum_{k'=k_0}^{k-2} \mathbb{E}\left(h(\mathsf{MC}_{k-1})\mathbb{1}_{\{\mathsf{MC}_k < \mathsf{MC}_{k-1}\}}\mathbb{1}_{\left\{\mathsf{MC}_{k-1} = \mathsf{MC}_{k'+1} < \mathsf{MC}_{k'}\right\}}\,\Big|\,\mathsf{MC}_{k_0} = m\right)$$
$$= \sum_{k'=k_0}^{k-2} \mathbb{E}\left(h(\mathsf{MC}_{k-1})\mathbb{P}\left(\mathsf{MC}_k < \mathsf{MC}_{k-1}\,\big|\,\mathsf{MC}_{k-1}\right)\mathbb{1}_{\left\{\mathsf{MC}_{k-1} = \mathsf{MC}_{k'+1} < \mathsf{MC}_{k'}\right\}}\,\Big|\,\mathsf{MC}_{k_0} = m\right)$$
$$= \sum_{k'=k_0}^{k-2} \mathbb{E}\left(h(\mathsf{MC}_{k-1})\left[1 - \mathsf{MC}_{k-1}(1 - \theta(\mathsf{MC}_{k-1}))\right]\mathbb{1}_{\left\{\mathsf{MC}_{k-1} = \mathsf{MC}_{k'+1} < \mathsf{MC}_{k'}\right\}}\,\Big|\,\mathsf{MC}_{k_0} = m\right)$$
$$= \sum_{k'=k_0}^{k-2} \mathbb{E}\left(h(\mathsf{MC}_{k'+1})\left[1 - \mathsf{MC}_{k'+1}(1 - \theta(\mathsf{MC}_{k'+1}))\right]\right.$$
$$\left. \cdot\,\mathbb{P}\left(\mathsf{MC}_{k-1} = \mathsf{MC}_{k'+1}\,\big|\,\mathsf{MC}_{k'+1}\right)\mathbb{1}_{\left\{\mathsf{MC}_{k'+1} < \mathsf{MC}_{k'}\right\}}\,\Big|\,\mathsf{MC}_{k_0} = m\right)$$
$$= \sum_{k'=k_0}^{k-2} \mathbb{E}\left(\int_1^{\mathsf{MC}_{k'}} \theta'(\tilde{m})d\tilde{m}\,\frac{1 - \mathsf{MC}_{k'}(1 - \theta(\mathsf{MC}_{k'}))}{\theta(\mathsf{MC}_{k'})}h(\tilde{m})\right.$$
$$\left. \cdot\,\left[1 - \tilde{m}(1 - \theta(\tilde{m}))\right]\left[\tilde{m}(1 - \theta(\tilde{m}))\right]^{k-k'-2}\,\Big|\,\mathsf{MC}_{k_0} = m\right) \tag{7.6}$$

by Proposition 7.1 (b). The fraction $(1 - x(1 - \theta(x)))/\theta(x)$ can be rewritten as $1 - (x-1)(1 - \theta(x))/\theta(x)$, which is clearly at most 1 for $x > 1$, and we can bound $\mathsf{MC}_{k'} \le \mathsf{MC}_{k_0} = m$.

Inserting this bound, we obtain

$$
\mathbb{E}\left(h(X_k^{\mathsf{BB}})\mathbb{1}_{\{\mathsf{MC}_k < \mathsf{MC}_{k-1} < \mathsf{MC}_{k_0}\}} \,\middle|\, \mathsf{MC}_{k_0} = m\right)
$$
$$
\leq \int_1^m \theta'(\tilde{m})h(\tilde{m})d\tilde{m} \sum_{k'=k_0}^{k-2} \left[1 - \tilde{m}(1-\theta(\tilde{m}))\right]\left[\tilde{m}(1-\theta(\tilde{m}))\right]^{k-k'-2}
$$
$$
\leq \int_1^m \theta'(\tilde{m})h(\tilde{m})d\tilde{m} \tag{7.7}
$$

by bounding the geometric series. By Proposition 6.1 (c) this completes the proof. $\quad\square$

7.2 First passage times and the IP backbone

In this section, we study $T_{\mathsf{BB}\,k}$, the time at which the kth backbone vertex is found. From now until the end of Section 7, we suppose that $(f_n(x))_n$ and $(s_n)_n$ satisfy Condition 2.2. We do not assume Condition 2.3 except where specifically noted. Write $y^+ = \max(y,0)$.

Lemma 7.3. *Given $m_0 \in (1,\infty)$, there is a constant $K < \infty$ such that*

$$
\mathbb{E}\left((T_{\mathsf{BB}\,k_1} - T_{\mathsf{BB}\,k_0} - f_n(m))^+ \,\middle|\, \mathsf{MC}_{k_0} = m\right) \leq \frac{K(k_1 - k_0)f_n(m)}{s_n} \tag{7.8}
$$

whenever $m \in (1, m_0]$ and $k_0, k_1 \in \mathbb{N}$ with $k_0 \leq k_1$.

Proof. The backbone edge weights $X_{k_0+1}^{\mathsf{BB}}, \ldots, X_{k_1}^{\mathsf{BB}}$ include at most one edge weight equal to MC_{k_0}. Consequently, $(T_{\mathsf{BB}\,k_1} - T_{\mathsf{BB}\,k_0} - f_n(m))^+ \leq \sum_{k=k_0+1}^{k_1} f_n(X_k^{\mathsf{BB}})\mathbb{1}_{\{X_k^{\mathsf{BB}} < \mathsf{MC}_{k_0}\}}$, so that

$$
\mathbb{E}\left((T_{\mathsf{BB}\,k_1} - T_{\mathsf{BB}\,k_0} - f_n(m))^+ \,\middle|\, \mathsf{MC}_{k_0} = m\right)
$$
$$
\leq \sum_{k=k_0+1}^{k_1} \mathbb{E}\left(f_n(X_k^{\mathsf{BB}})\mathbb{1}_{\{X_k^{\mathsf{BB}} < \mathsf{MC}_{k_0}\}} \,\middle|\, \mathsf{MC}_{k_0} = m\right)
$$
$$
\leq \sum_{k=k_0+1}^{k_1} C \int_0^m f_n(x)dx \tag{7.9}
$$

by Lemma 7.2. Lemma 5.3 completes the proof. $\quad\square$

Corollary 7.4. *If $\sigma_n = O(s_n)$ then*

$$
f_n^{-1}(T_{\mathsf{BB}\,\sigma_n}) \leq \mathsf{MC}_0 + O_{\mathbb{P}}(1/s_n) \qquad \text{as } n \to \infty. \tag{7.10}
$$

Proof. Apply Lemma 7.3 with $k_1 = \lceil \sigma_n \rceil$ and $k_0 = 0$ (so that $T_{\mathsf{BB}\,k_0} = 0$) to find that

$$
\mathbb{E}\left(T_{\mathsf{BB}\,\sigma_n} \,|\, \mathsf{MC}_0 = m\right) \leq f_n(m) + O(f_n(m)) \tag{7.11}
$$

for $m \in (1, m_0]$. In particular, $T_{\mathsf{BB}\,\sigma_n} \leq O_{\mathbb{P}}(f_n(\mathsf{MC}_0))$ conditional on $\{\mathsf{MC}_0 \in (1, m_0]\}$. By Lemma 5.1 with $y_0 = f_n(m)$ and $y_1 = T_{\mathsf{BB}\sigma_n}$, we conclude that (7.10) holds conditional on $\{\mathsf{MC}_0 \in (1, m_0]\}$. Since MC_0 does not depend on n, we can make $\mathbb{P}(\mathsf{MC}_0 \in (1, m_0])$ arbitrarily close to 1 by taking m_0 large enough, and it follows that (7.10) holds unconditionally. $\quad\square$

Corollary 7.5. *If $\sigma_n \to \infty$ and $\sigma_n = O(s_n)$, then*

$$
f_n^{-1}(T_{\mathsf{BB}\,\sigma_n}) \xrightarrow{\mathbb{P}} \mathsf{MC}_0 \qquad \text{as } n \to \infty. \tag{7.12}
$$

Proof. In view of Corollary 7.4, it suffices to prove the matching lower bound $T_{\mathrm{BB}\,\sigma_n} \geq f_n(\mathrm{MC}_0)$ whp. It suffices to show that $\sigma_n \geq I_0$ whp, where I_0 is the random height at which the maximum defining MC_0 is attained; in this case, $T_{\mathrm{BB}\,\sigma_n}$ will be a sum of FPP edge weights one of which is $f_n(\mathrm{MC}_0)$. But by Proposition 7.1 (a) I_0 is finite a.s. and does not depend on n, whereas $\sigma_n \to \infty$ by assumption. $\qquad\square$

7.3 First passage times and the IP cluster: proof of Theorem 2.16

In this section, we investigate how many children of the off-backbone tree τ_k are invaded before time $T_{\mathrm{BB}k_1}$. Further, we study the length of the backbone at the time $T_{\mathrm{noninvaded}}$ (recall (2.27)) at which the first difference appears between the branching process and IP dynamics, as well as various estimates on lucky vertices.

Given $k \in \mathbb{N} \cup \{0\}$, $k_1 \in \mathbb{N}$ with $k < k_1$, define

$$N(k, k_1) = |\{v \in \partial\tau_k \colon T_v \leq T_{\mathrm{BB}\,k_1}\}|. \tag{7.13}$$

That is, $N(k, k_1)$ is the number of uninvaded children $v \notin \tau_k$ of invaded vertices from τ_k that are explored before the backbone vertex at height k_1.

Lemma 7.6. *Let $m_0 \in (1, \infty)$ be given. Then there exists a constant $C < \infty$ such that*

$$\mathbb{E}\left(N(k, k_1) \,\middle|\, (\mathrm{MC}_{k'}, \tau_{k'})_{k'=0}^{\infty}\right) \leq \frac{C(k_1 - k)\,|\tau_k|}{s_n^2} \qquad \text{on } \{T_{\mathrm{BB}\,k_1} \leq f_n(m_0)\}. \tag{7.14}$$

Proof. Define

$$\tilde{X}(k, k_1) = f_n^{-1}(T_{\mathrm{BB}\,k_1} - T_{\mathrm{BB}\,k}) = \sup\{x \colon f_n(x) \leq T_{\mathrm{BB}\,k_1} - T_{\mathrm{BB}\,k}\}. \tag{7.15}$$

By construction, any descendant of V_k^{BB} that is explored by time $T_{\mathrm{BB}\,k_1}$ must be connected to V_k^{BB} by a path whose PWIT edge weights are at most $\tilde{X}(k, k_1)$. In particular, $N(k, k_1)$ is bounded above by the number of vertices $w \in \partial\tau_k$ with edge weights X_w in the interval $(\mathrm{MC}_k, \tilde{X}(k, k_1))$. By Proposition 7.1 (g), the number of such vertices has a Poisson distribution, and summing over the choice of parent $v = p(w) \in \tau_k$ gives

$$\mathbb{E}\left(N(k, k_1) \,\middle|\, (\mathrm{MC}_{k'}, \tau_{k'})_{k'=0}^{\infty}\right) \leq |\tau_k|\,\mathbb{E}\left(\left(\tilde{X}(k, k_1) - \mathrm{MC}_k\right)^+ \,\middle|\, (\mathrm{MC}_{k'}, \tau_{k'})_{k'=0}^{\infty}\right). \tag{7.16}$$

Applying Lemma 5.1 and Lemma 7.3 yields

$$\begin{aligned}
\mathbb{E}&\left(N(k, k_1) \,\middle|\, (\mathrm{MC}_{k'}, \tau_{k'})_{k'=0}^{\infty}\right) \\
&\leq |\tau_k|\,\mathbb{E}\left(\frac{K}{s_n f_n(\mathrm{MC}_k)}\left(f_n(\tilde{X}(k, k_1)) - f_n(\mathrm{MC}_k)\right)^+ \,\middle|\, (\mathrm{MC}_{k'}, \tau_{k'})_{k'=0}^{\infty}\right) \\
&\leq \frac{K\,|\tau_k|}{s_n f_n(\mathrm{MC}_k)}\,\mathbb{E}\left(\left(T_{\mathrm{BB}\,k_1} - T_{\mathrm{BB}\,k} - f_n(\mathrm{MC}_k)\right)^+ \,\middle|\, (\mathrm{MC}_{k'}, \tau_{k'})_{k'=0}^{\infty}\right) \\
&\leq \frac{K\,|\tau_k|}{s_n f_n(\mathrm{MC}_k)}\,\frac{K'(k_1 - k)f_n(\mathrm{MC}_k)}{s_n} \qquad \text{on } \{T_{\mathrm{BB}\,k_1} \leq f_n(m_0)\}. \qquad\square
\end{aligned}$$

Lemma 7.7. *Let $R_{\mathrm{noninvaded}} = \min\{k \colon T_{\mathrm{BB}\,k} > T_{\mathrm{noninvaded}}\}$, the height of the lowest unexplored backbone vertex by time $T_{\mathrm{noninvaded}}$. Then $R_{\mathrm{noninvaded}}$ is at least of order $s_n^{2/3}$, i.e., $R_{\mathrm{noninvaded}}^{-1} = O_{\mathbb{P}}(s_n^{-2/3})$.*

Proof. It suffices to show that if $(\sigma_n)_n$ is any sequence satisfying $\sigma_n = o(s_n^{2/3})$, then $R_{\mathrm{noninvaded}} > \sigma_n$ whp. The latter event can be written as $\{N(k, \sigma_n) = 0$ for all $k = 0, 1, \ldots, \sigma_n - 1\}$: for if, to the contrary, the first non-invaded vertex is born before time $T_{\mathrm{BB}\,\sigma_n}$, it must have a parent in some τ_k with $k < \sigma_n$.

Let $\varepsilon > 0$ be given. By Corollary 7.5 and the fact that MC_0 does not depend on n, we may choose m_0 so that $\mathbb{P}(T_{\mathsf{BB}\,\sigma_n} \le m_0) \ge 1 - \varepsilon$ for all n. Applying Markov's inequality and Lemma 7.6,

$$
\begin{aligned}
\mathbb{P}\left(R_{\mathrm{noninvaded}} \le \sigma_n \,|\, (\mathsf{MC}_{k'}, \tau_{k'})_{k'=0}^{\infty} \right) &= \mathbb{P}\left(\sum_{k=0}^{\sigma_n - 1} N(k, \sigma_n) \ge 1 \,\bigg|\, (\mathsf{MC}_{k'}, \tau_{k'})_{k'=0}^{\infty} \right) \\
&\le \sum_{k=0}^{\sigma_n - 1} \mathbb{E}\left(N(k, \sigma_n) \,|\, (\mathsf{MC}_{k'}, \tau_{k'})_{k'=0}^{\infty} \right) \\
&\le \frac{C\sigma_n}{s_n^2} \sum_{k=0}^{\sigma_n - 1} |\tau_k|
\end{aligned}
\tag{7.17}
$$

on $\{T_{\mathsf{BB}\,\sigma_n} \le m_0\}$. The last sum in (7.17) is the size of that part of $\mathrm{IP}(\infty)$ not descended from $V_{\sigma_n}^{\mathsf{BB}}$, and is therefore $O_{\mathbb{P}}(\sigma_n^2)$ by Proposition 7.1 (i). Consequently the upper bound in (7.17) is $O_{\mathbb{P}}(\sigma_n^3/s_n^2)$ and is therefore $o_{\mathbb{P}}(1)$ by assumption. $\qquad\square$

To complete the proof of Theorem 2.16, we will show that $\mathsf{BP}_{\mathsf{BB}\,K\sigma_n}$ is likely to contain a σ_n^2-lucky backbone vertex if K is sufficiently large, uniformly in the choice of $(\sigma_n)_n$. In order to achieve (conditional) independence between different backbone vertices (which are randomly chosen and impose a conditioning on their neighbouring edges) we strengthen the definition of lucky vertex. We say that a backbone vertex X_k^{BB} is *q-backbone-lucky* if $\left|\tau_k \cap \mathsf{BP}_{T_{\mathsf{BB}\,k}+f_n(1)}\right| \ge q$ – i.e., if X_k^{BB} has at least q descendants by age $f_n(1)$ when we exclude descendants of X_{k+1}^{BB}. Evidently, a q-backbone-lucky vertex is also q-lucky.

Lemma 7.8. *Conditional on the backbone vertices and weights $(\mathsf{MC}_{k'}, V_{k'}^{\mathsf{BB}}, X_{k'}^{\mathsf{BB}})_{k'=0}^{\infty}$, the probability of being q-backbone-lucky is*

$$
\begin{aligned}
\mathbb{P}\left(X_k^{\mathsf{BB}} \text{ is q-backbone-lucky} \,\big|\, (\mathsf{MC}_{k'}, V_{k'}^{\mathsf{BB}}, X_{k'}^{\mathsf{BB}}), \mathsf{MC}_k = m \right) \\
= \mathbb{P}\left(\left|\mathsf{BP}_{f_n(1)}\right| \ge q \,\big|\, |\mathsf{PGW}_m| < \infty \right), \quad (7.18)
\end{aligned}
$$

the conditional probability from Lemma 6.3. Moreover, for any choice of natural numbers $(q_k)_k$, the events $\left\{ X_k^{\mathsf{BB}} \text{ is } q_k\text{-backbone-lucky} \right\}$, $k \ge 0$, are conditionally independent.

Proof. This follows immediately from the fact that, conditionally on $\{\mathsf{MC}_k = m\}$, the subtree τ_k equipped with its edge weights (considered relative to V_k^{BB}) has the same distribution as the PWIT (considered relative to the root $\varnothing_1$) under the conditioning $|\mathsf{PGW}_m| < \infty$. See Proposition 7.1 (d) and (f) and Proposition 6.1 (d). $\qquad\square$

Lemma 7.9. *Let $\varepsilon > 0$ and $K < \infty$ be given. Then there exists $K' < \infty$ such that, for any sequence $(\sigma_n)_n$ with $1 \le \sigma_n \le Ks_n$,*

$$
\mathbb{P}(T_{\sigma_n^2\text{-lucky}} \le T_{\mathsf{BB}\,K'\sigma_n}) \ge 1 - \varepsilon.
\tag{7.19}
$$

In particular, since $T_{k\text{-lucky}} \ge T_{\mathrm{size}\,k}$, Lemma 7.9 shows that $|\mathsf{BP}_{\mathsf{BB}\,\sigma_n}|$ has size of order at least σ_n^2.

Proof. By Proposition 7.1 (c), there is a constant $C < \infty$ such that

$$
\mathbb{P}(\mathsf{MC}_{\lceil\sigma_n\rceil} \le 1 + C/\sigma_n) \ge 1 - \tfrac{1}{2}\varepsilon
\tag{7.20}
$$

for any choice of $\sigma_n \ge 1$. Apply Lemmas 7.8 and 6.3 (with $m_0 = 1 + C$ and with the constant $K_{6.3}$ from that lemma chosen large enough so that the conclusion applies with

k replaced by $\lceil \sigma_n^2 \rceil$ and $m \in [1, 1 + C/\sigma_n]$; we may choose such a constant because $\lceil \sigma_n^2 \rceil \leq 2\sigma_n^2 \leq 2K s_n^2$ and $|m - 1|^{-2} \geq \sigma_n^2/C^2$) to find

$$
\begin{aligned}
&\mathbb{P}\left(T_{\sigma_n^2\text{-lucky}} \leq T_{\mathsf{BB}\,K'\sigma_n} \,\middle|\, \mathsf{MC}_{\lceil \sigma_n \rceil} \leq 1 + C/\sigma_n\right) \\
&\quad \geq \mathbb{P}\left(\text{some } X_k^{\mathsf{BB}},\, \sigma_n \leq k \leq K'\sigma_n,\, \text{is } \lceil \sigma_n^2 \rceil\text{-backbone-lucky} \,\middle|\, \mathsf{MC}_{\lceil \sigma_n \rceil} \leq 1 + C/\sigma_n\right) \\
&\quad \geq 1 - \left(1 - \frac{\delta}{\sqrt{\lceil \sigma_n^2 \rceil}}\right)^{\lfloor K'\sigma_n \rfloor - \lceil \sigma_n \rceil}.
\end{aligned}
\tag{7.21}
$$

The lower bound in (7.21) can be made at least $1 - \frac{1}{2}\varepsilon$ by choosing K' large enough, uniformly over $\sigma_n \geq 1$. Together with (7.20), this completes the proof. $\qquad\square$

Proof of Theorem 2.16. As in the proof of Lemma 7.7, it suffices to show that if $(\sigma_n)_n$ is any sequence satisfying $\sigma_n = o(s_n^{2/3})$, then $|\mathsf{BP}_{\text{noninvaded}}| > \sigma_n^2$ whp, or equivalently $T_{\text{size}\,\sigma_n^2} < T_{\text{noninvaded}}$ whp.

Let $\varepsilon > 0$ be given. Applying Lemma 7.9 (with $K = 1$ and with a term $o(1)$ to account for the fact that $\sigma_n \leq s_n$ might fail for small n),

$$
\begin{aligned}
\mathbb{P}\left(T_{\text{size}\,\sigma_n^2} < T_{\text{noninvaded}}\right) &\geq \mathbb{P}\left(T_{\mathsf{BB}\,K'\sigma_n} < T_{\text{noninvaded}}\right) - \varepsilon - o(1) \\
&= \mathbb{P}\left(R_{\text{noninvaded}} > K'\sigma_n\right) - \varepsilon - o(1).
\end{aligned}
\tag{7.22}
$$

Since $K'\sigma_n$ is still $o(s_n^{2/3})$, the last probability tends to 1 by Lemma 7.7. Since $\varepsilon > 0$ was arbitrary, this completes the proof. $\qquad\square$

7.4 Proof of Theorem 2.15

In this section, we prove Theorem 2.15, which yields detailed estimates on the times when the sizes and heights of the branching process tree reach σ_n^2 and σ_n, respectively, as well as the times where the backbone contains σ_n vertices and the first σ_n^2-lucky vertex is found. Further, the sizes and heights of the branching process tree at those times is bounded. This is achieved by using appropriate couplings to Poisson branching processes. The first key step will be to show that $\mathsf{BP}_{\mathsf{BB}\,\sigma_n}$ contains $O_{\mathbb{P}}(\sigma_n^2)$ vertices and has diameter $O_{\mathbb{P}}(\sigma_n)$, as formulated in Lemma 7.12. We start by proving some preliminary estimates that build up towards the proof of Lemma 7.12. After that, we complete the proof of Theorem 2.15.

Define

$$
\begin{aligned}
B(k, k_1) &= \left\{v \in \partial\tau_k \colon T_v \leq T_{\mathsf{BB}\,k_1}\right\}, \\
D^{(v)}(k, k_1) &= \left\{w \colon T_w \leq T_{\mathsf{BB}\,k_1} \text{ and } p^\ell(w) = v \text{ for some } \ell \geq 0\right\}, \quad (v \in B(k, k_1)).
\end{aligned}
\tag{7.23}
$$

Thus, the random variable $N(k, k_1)$ from (7.13) equals $|B(k, k_1)|$, and $D^{(v)}(k, k_1)$ is the subtree of descendants of $v \in B(k, k_1)$ explored by time $T_{\mathsf{BB}\,k_1}$. Write $D^{(v, \geq r)}(k, k_1)$ (respectively, $D^{(v, \leq r)}(k, k_1)$) for the parts of the tree $D^{(v)}(k, k_1)$ at least r generations away from v (respectively, within r generations of v). Define

$$
\begin{aligned}
N^{(\geq r)}(k, k_1) &= \left|\left\{v \in B(k, k_1) \colon D^{(v, \geq r)}(k, k_1) \neq \varnothing\right\}\right|, \\
U^{(\leq r)}(k, k_1) &= \sum_{v \in B(k, k_1)} \left|D^{(v, \leq r)}(k, k_1)\right|,
\end{aligned}
\tag{7.24}
$$

i.e., $N^{(\geq r)}(k, k_1)$ is the number of vertices from $\partial\tau_k$ with at least one descendant at distance at least r that is explored before time $T_{\mathsf{BB}\,k_1}$, and $U^{(\leq r)}(k, k_1)$ is the total number, among vertices $v \in B(k, k_1)$ and their descendants within at most r generations, that are explored before time $T_{\mathsf{BB}\,k_1}$. (Thus $N^{(\geq r)}(k, k_1)$ reduces to $N(k, k_1)$ when $r = 0$.) To

bound the diameter of $\mathsf{BP}_{\mathsf{BB}\,\sigma_n}$, we will show that if r is sufficiently large compared to σ_n, then $N^{(\geq r)}(k,\sigma_n) = 0$ for all $k < \sigma_n$ with probability close to 1. Having done so, we will then show that $\sum_{k<\sigma_n} U^{(\leq r)}(k,\sigma_n) = O_{\mathbb{P}}(\sigma_n^2)$ using moment bounds.

These estimates are made by comparing $D^{(v)}(k,k_1)$ to appropriate Poisson Galton–Watson branching trees. For $m \geq 0$ and $v \in \mathcal{T}$, define

$$\mathsf{PGW}_m^{(v)} = \left\{ w \in \mathcal{T} \colon p^\ell(w) = v \text{ for some } \ell \text{ and } X_{p^j(w)} \leq m \text{ for } j = 0, \ldots, \ell \right\}, \qquad (7.25)$$

i.e., the subtree of descendants of v connected to v by a path of PWIT edge weights at most m. (Thus Proposition 7.1 (h) implies that, other than being rooted at v instead of $\varnothing_1$, the trees $\mathsf{PGW}_m^{(v)}$, $v \in B(k,k_1)$, are i.i.d. Poisson Galton–Watson trees with mean m, conditionally independent over the choice of v and k.) The key observation for our purposes is that

$$D^{(v)}(k,k_1) \subset \mathsf{PGW}_{\tilde{X}(k,k_1)}^{(v)} \qquad (v \in B(k,k_1)) \qquad (7.26)$$

where $\tilde{X}(k,k_1)$ was defined in (7.15).

We begin by showing that whp, all the Poisson Galton–Watson trees $\mathsf{PGW}^{(v)}$, $v \in B(k,k_1)$, remain finite until we reach diameter of order s_n:

Lemma 7.10. *Let $\varepsilon > 0$ and $K < \infty$ be given. Then there exists $\delta > 0$ such that*

$$\mathbb{P}\left(\left| \mathsf{PGW}_{\tilde{X}(k,Ks_n)}^{(v)} \right| < \infty \text{ for all } v \in B(k,Ks_n) \text{ and for all } k \leq \delta s_n \right) \geq 1 - \varepsilon \qquad (7.27)$$

for all n sufficiently large.

Proof. For notational convenience, consider a fixed $k < Ks_n$ and abbreviate $\tilde{X}(k,Ks_n)$ as $\tilde{X}$. Denote

$$\tilde{N} = \left| \left\{ v \in B(k,Ks_n) \colon \left| \mathsf{PGW}_{\tilde{X}}^{(v)} \right| < \infty \right\} \right|. \qquad (7.28)$$

Similar to the proof of Lemma 7.6, we use moment bounds on $\tilde{N}$; this time, the bounds will be essentially uniform over k.

By Proposition 7.1 (h), each of the $N(k,Ks_n)$ vertices $v \in B(k,Ks_n)$ has probability $\theta(\tilde{X})$ of surviving forever, and therefore

$$\mathbb{E}\left(\tilde{N} \,\middle|\, (\mathsf{MC}_{k'}, X_{k'}, \tau_{k'}) \right) = \theta(\tilde{X}) \mathbb{E}\left(N(k,Ks_n) \,\middle|\, (\mathsf{MC}_{k'}, X_{k'}, \tau_{k'}) \right)$$
$$\leq C(\tilde{X}-1)^+ (\tilde{X} - \mathsf{MC}_k)^+ |\tau_k|$$
$$= C\left((\tilde{X} - \mathsf{MC}_k)^+ \right)^2 |\tau_k| + C(\tilde{X} - \mathsf{MC}_k)^+ (\mathsf{MC}_k - 1) |\tau_k|, \quad (7.29)$$

where we used $\theta(x) \leq C(x-1)^+$ by Proposition 6.1 (b).

By Corollary 7.5, we may choose $m_0 \in (1,\infty)$ so that the event $\{\mathsf{MC}_0 \leq m_0, T_{\mathsf{BB}\,Ks_n} \leq f_n(m_0)\}$ has probability at least $1 - \varepsilon/4$. In the remainder of the proof we assume that this event occurs.

To bound the second term in (7.29), recall from Proposition 7.1 (d) that, conditionally on MC_k, $|\tau_k|$ is the total progeny of a Poisson Galton–Watson branching process with mean $\widehat{\mathsf{MC}}_k$. In particular,

$$\mathbb{E}\left(|\tau_k| \,\middle|\, \mathsf{MC}_k \right) = \frac{1}{1 - \widehat{\mathsf{MC}}_k}. \qquad (7.30)$$

Since $1 - \hat{m} \sim m - 1$ as $m \downarrow 1$ (see Proposition 6.1 (d)) it follows that

$$\mathbb{E}\left((\mathsf{MC}_k - 1) |\tau_k| \,\middle|\, \mathsf{MC}_k \right) \leq C' \quad \text{on } \{\mathsf{MC}_k \leq m_0\} \qquad (7.31)$$

for some $C' < \infty$. As in the proof of Lemma 7.6, we can apply Lemmas 5.1 and 7.3 to conclude that

$$\mathbb{E}\left(C(\tilde{X} - \mathsf{MC}_k)(\mathsf{MC}_k - 1)\,|\tau_k|\,\big|\,\mathsf{MC}_k\right) \leq \frac{C''(Ks_n)}{s_n^2} \quad \text{on } \{\mathsf{MC}_0 \leq m_0, T_{\mathsf{BB}\,Ks_n} \leq f_n(m_0)\}. \tag{7.32}$$

To bound the first term in (7.29), use instead Lemmas 5.2 and 7.3 to conclude that

$$\mathbb{E}\left(C((\tilde{X} - \mathsf{MC}_k)^+)^2\,|\tau_k|\,\big|\,\mathsf{MC}_k, \tau_k\right) \leq \frac{K'}{s_n^2}\frac{\mathbb{E}\left((T_{\mathsf{BB}\,Ks_n} - T_{\mathsf{BB}\,k} - f_n(\mathsf{MC}_k))^+\,|\,\mathsf{MC}_k\right)}{f_n(\mathsf{MC}_k)}|\tau_k|$$

$$\leq \frac{K''s_n\,|\tau_k|}{s_n^3} = \frac{K''\,|\tau_k|}{s_n^2}. \tag{7.33}$$

By Proposition 7.1 (i), there is a constant C''' such that $\mathbb{P}(\sum_{k \leq \delta s_n} \tau_k \leq C'''\lceil \delta s_n\rceil^2) \geq 1 - \varepsilon/4$ for all $\delta > 0$. Assuming that this event occurs, we may sum (7.32)–(7.33) over $k \leq \delta s_n$ to obtain

$$\mathbb{E}\left(\tilde{N}\mathbb{1}_A\right) \leq C''''\left(\frac{\lceil \delta s_n\rceil^2}{s_n^2} + \frac{\delta s_n + 1}{s_n}\right), \tag{7.34}$$

where $A = \left\{\mathsf{MC}_0 \leq m_0, T_{\mathsf{BB}\,Ks_n} \leq f_n(m_0), \sum_{k \leq \delta s_n} \tau_k \leq C'''\lceil \delta s_n\rceil^2\right\}$ is an event of probability at least $1 - \varepsilon/2$. Since $s_n \to \infty$, we can make the upper bound smaller than $\varepsilon/2$ for sufficiently large n by taking $\delta > 0$ small enough. By Markov's inequality, this completes the proof. $\qquad\square$

Lemma 7.10 shows that the trees $\mathsf{PGW}^{(v)}_{\tilde{X}(k,Ks_n)}$ can be assumed to be finite for $v \in B(k, Ks_n)$, $k \leq \delta s_n$. For such k, we may use Proposition 6.1 (d) and facts about critical Poisson Galton–Watson trees to bound the size and diameter of the subtrees $D^{(v)}(k, Ks_n)$. For larger k, Proposition 7.1 (c) and Lemma 7.3 imply that MC_k and $\tilde{X}(k, Ks_n)$ are close to 1, and we can bound the probability of a large diameter in $D^{(v)}(k, Ks_n)$ in terms of critical Poisson Galton–Watson trees.

For $0 < x < 1 \leq m$ and $\ell \in \mathbb{N}$, let $P^{(v)}(m, \ell, x)$ denote the subtree of $\mathsf{PGW}^{(v)}_m$ consisting of descendants w connected to v by a path containing at most ℓ PWIT edge weights in the interval $(x, m]$. Write $P^{(v, \geq r)}(m, \ell, x)$ for the subtree of $P^{(v)}(m, \ell, x)$ within distance r of v.

Lemma 7.11. *There is a constant C such that*

$$\mathbb{P}\left(|P^{(v, \geq r)}(m, \ell, x)| > 0\right) \leq \frac{C}{r}\left(\frac{m-x}{1-x}\right)^\ell \tag{7.35}$$

for all $0 < x < 1 \leq m$ and $\ell \in \mathbb{N}$.

Proof. On the event $\{|P^{(v, \geq r)}(m, \ell, x)| > 0\}$, let W be chosen uniformly from the vertices of $P^{(v, \geq r)}(m, \ell, x)$ at distance exactly r from v. Let L denote the collection of vertices w along the path from v to W for which $x < X_w \leq m$, so that $|L| \leq \ell$ by definition. By the Poisson point process property, conditionally on the occurrence of $\{|P^{(v, \geq r)}(m, \ell, x)| > 0\}$ and the values of W and L, the edge weights X_w, $w \in L$, are uniformly distributed on $(x, m]$. In particular, $x < X_w \leq 1$ for each $w \in L$ with conditional probability $[(1-x)/(m-x)]^{|L|}$, which is at least $[(1-x)/(m-x)]^\ell$ by construction. If this additional event occurs then $\left|\mathsf{PGW}^{(\geq r)}_1\right| > 0$, so

$$\mathbb{P}\left(|P^{(v, \geq r)}(m, \ell, x)| > 0\right) \cdot [(1-x)/(m-x)]^\ell \leq \mathbb{P}\left(\left|\mathsf{PGW}^{(\geq r)}_1\right| > 0\right) \tag{7.36}$$

and (7.35) follows by the bound $\mathbb{P}\left(\left|\mathsf{PGW}^{(\geq r)}_1\right| > 0\right) = O(1/r)$ from Proposition 6.1 (h). $\quad\square$

Lemma 7.12. *Let $K < \infty$ be given. Then $\mathsf{BP}_{\mathsf{BB}\,\sigma_n}$ contains $O_\mathbb{P}(\sigma_n^2)$ vertices and has diameter $O_\mathbb{P}(\sigma_n)$, uniformly over the choice of sequence $(\sigma_n)_n$ with $1 \leq \sigma_n \leq Ks_n$.*

Proof. Let $K < \infty$ be given. By Proposition 7.1 (i), the part of $\mathrm{IP}(\infty)$ not descended from $V^{\mathrm{BB}}_{\lceil \sigma_n \rceil}$ has size $O_\mathbb{P}(\sigma_n^2)$ and diameter $O_\mathbb{P}(\sigma_n)$, uniformly over the choice of $(\sigma_n)_n$. So it suffices to show that the non-invaded subtrees $D^{(v)}(k, \sigma_n)$, $v \in B(k, \sigma_n)$, $k \leq \sigma_n$, have total size $O_\mathbb{P}(\sigma_n^2)$ and maximum diameter $O_\mathbb{P}(\sigma_n)$, uniformly over the choice of $(\sigma_n)_n$ with $\sigma_n \leq K s_n$.

Let $\varepsilon > 0$ be given and choose $m_0 \in (1, \infty)$ such that

$$\mathbb{P}(M_0 \leq m_0, T_{\mathrm{BB}\, K s_n} \leq f_n(m_0)) \geq 1 - \varepsilon. \tag{7.37}$$

We assume that this event occurs for the rest of the proof.

Let $\delta > 0$ be the constant from Lemma 7.10. We consider first $k \leq \delta s_n$. Since $D^{(v)}(k, \sigma_n) \subset D^{(v)}(k, K s_n)$ for any k, it will suffice to consider the case $\left|\mathrm{PGW}^{(v)}_{\tilde{X}(k, K s_n)}\right| < \infty$ for $k \leq \delta s_n$. Define

$$N^{(\geq r, \mathrm{finite})}(k, k_1) = \left| \left\{ v \in B(k, k_1) \colon D^{(v, \geq r)}(k, k_1) \neq \varnothing \text{ and } \left|\mathrm{PGW}^{(v)}_{\tilde{X}(k, K s_n)}\right| < \infty \right\} \right|,$$
$$U^{(\leq r, \mathrm{finite})}(k, k_1) = \sum_{v \in B(k, k_1)} \left| D^{(v, \leq r)}(k, k_1) \right| \mathbb{1}_{\left\{ \left|\mathrm{PGW}^{(v)}_{\tilde{X}(k, K s_n)}\right| < \infty \right\}}. \tag{7.38}$$

By (7.26), Proposition 6.1 (d) and (h), and Lemma 7.6,

$$\mathbb{E}\left(N^{(\geq r, \mathrm{finite})}(k, K s_n) \,\middle|\, (\mathsf{MC}_{k'}, \tau_{k'}) \right)$$
$$\leq \mathbb{E}\left(N(k, K s_n) \,\middle|\, (\mathsf{MC}_{k'}, \tau_{k'}) \right) \mathbb{P}\left(\mathrm{PGW}^{(v, \geq r)}_{\tilde{X}(k, K s_n)} \neq \varnothing, \left|\mathrm{PGW}^{(v)}_{\tilde{X}(k, K s_n)}\right| < \infty \right)$$
$$\leq \frac{C(K s_n) \left|\tau_k\right|}{s_n^2} \mathbb{P}\left(\mathrm{PGW}^{(v, \geq r)}_{\tilde{X}(k, K s_n)} \neq \varnothing \,\middle|\, \left|\mathrm{PGW}^{(v)}_{\tilde{X}(k, K s_n)}\right| < \infty \right)$$
$$\leq \frac{C(K s_n) \left|\tau_k\right|}{s_n^2} \mathbb{P}\left(\mathrm{PGW}^{(v, \geq r)}_1 \neq \varnothing \right) \leq \frac{C' \left|\tau_k\right|}{r s_n}. \tag{7.39}$$

Sum over $k \leq \sigma_n \wedge (\delta s_n)$ and use Proposition 7.1 (i) to find that

$$\mathbb{E}\left(\sum_{k \leq \sigma_n \wedge (\delta s_n)} N^{(\geq r, \mathrm{finite})}(k, K s_n) \,\middle|\, (\mathsf{MC}_{k'}, \tau_{k'}) \right)$$
$$\leq O(1/r s_n) \sum_{k \leq \sigma_n} |\tau_k| = O_\mathbb{P}\left(\frac{\sigma_n^2}{r s_n} \right) = O_\mathbb{P}\left(\frac{\sigma_n}{r} \right), \quad (7.40)$$

uniformly over the choice of the as-yet-unspecified constant r. Hence, by taking $r = C'' \sigma_n$ for C'' large enough, we may assume that

$$\mathbb{P}\left(N^{(\geq r, \mathrm{finite})}(k, K s_n) = 0 \text{ for all } k \leq \delta s_n \right) \geq 1 - 2\varepsilon. \tag{7.41}$$

With the same choice of r,

$$\mathbb{E}\left(U^{(\leq r, \mathrm{finite})}(k, K s_n) \,\middle|\, (\mathsf{MC}_{k'}, \tau_{k'}) \right)$$
$$\leq \mathbb{E}\left(N(k, K s_n) \,\middle|\, (\mathsf{MC}_{k'}, \tau_{k'}) \right) \mathbb{E}\left(\left|\mathrm{PGW}^{(v, \leq r)}_{\tilde{X}(k, K s_n)}\right| \mathbb{1}_{\left\{ \left|\mathrm{PGW}^{(v)}_{\tilde{X}(k, K s_n)}\right| < \infty \right\}} \right)$$
$$\leq \frac{C(K s_n) \left|\tau_k\right|}{s_n^2} \mathbb{E}\left(\left|\mathrm{PGW}^{(v, \leq r)}_{\tilde{X}(k, K s_n)}\right| \,\middle|\, \left|\mathrm{PGW}^{(v)}_{\tilde{X}(k, K s_n)}\right| < \infty \right)$$
$$\leq \frac{C(K s_n) \left|\tau_k\right|}{s_n^2} \mathbb{E}\left(\left|\mathrm{PGW}^{(v, \leq r)}_1\right| \right) \leq \frac{C(K s_n) \left|\tau_k\right|}{s_n^2} (r + 1), \tag{7.42}$$

since each generation of a critical Galton–Watson branching process has average size 1. Summing, we conclude similarly that

$$\sum_{k \leq \sigma_n \wedge (\delta s_n)} U^{(\leq r,\text{finite})}(k, K s_n) = O_{\mathbb{P}}\left(\frac{r \sigma_n^2}{s_n}\right) = O_{\mathbb{P}}\left(\sigma_n^2\right), \tag{7.43}$$

since $r = O(s_n)$.

We now turn to $k > \delta s_n$. Since k parametrizes height along the backbone, we only need to consider this range of k when $\sigma_n > \delta s_n$, passing, if necessary, to an appropriate subsequence of $(\sigma_n)_n$. In particular, it will suffice to prove that the subtrees $D^{(v)}(k, K s_n)$ have total size $O_{\mathbb{P}}(s_n^2)$ and maximum diameter $O_{\mathbb{P}}(s_n)$, since $s_n = O(\sigma_n)$ along the subsequence with $\sigma_n > \delta s_n$.

By Proposition 7.1 (c) and the fact that $k \mapsto \mathsf{MC}_k$ is non-increasing a.s., there exists $K' < \infty$ such that

$$\mathbb{P}\left(\mathsf{MC}_k \vee \tilde{X}(k, K s_n) \leq 1 + K'/s_n \text{ for all } k > \delta s_n\right) \geq 1 - \varepsilon. \tag{7.44}$$

Apply Condition 2.3 to find $\ell \in \mathbb{N}$ such that

$$f_n(1 + K'/s_n) \leq \ell f_n(1 - \eta/s_n) \tag{7.45}$$

for all n sufficiently large. In the rest of the proof, assume that the event from (7.44) occurs and that n is so large that (7.45) holds. Then $D^{(v)}(k, K s_n) \subset P^{(v)}(m, \ell, x)$ for all $k > \delta s_n$ with $m = 1 + K'/s_n$ and $x = 1 - \eta/s_n$. In particular, for this choice of m and x, the ratio $(m - x)/(1 - x)$ is bounded, so Lemma 7.11 implies that $\mathbb{P}(D^{(v, \geq r')}(k, K s_n) \neq \varnothing) \leq C'''/r'$ for some $C''' < \infty$. We can then repeat the argument from (7.39)–(7.40) to conclude that

$$\mathbb{E}\left(\sum_{\delta s_n < k \leq K s_n} N^{(\geq r')}(k, K s_n) \,\Big|\, (\mathsf{MC}_{k'}, \tau_{k'})\right) = O_{\mathbb{P}}\left(\frac{s_n}{r'}\right), \tag{7.46}$$

uniformly over the choice of r', and we may assume that

$$\mathbb{P}\left(N^{(\geq r')}(k, K s_n) = 0 \text{ for all } \delta s_n < k \leq K s_n\right) \geq 1 - 2\varepsilon \tag{7.47}$$

by taking $r' = C''''s_n$ for $C''''s_n$ sufficiently large. With this choice of r', note that $(1 + K'/s_n)^{r'}$ is bounded as $n \to \infty$. Since $\mathbb{E}\left(\left|\mathsf{PGW}^{(v, \leq r')}_{1+K'/s_n}\right|\right) \leq (r' + 1)(1 + K'/s_n)^{r'}$, we can repeat the argument from (7.42)–(7.43) to conclude that

$$\sum_{\delta s_n < k \leq K s_n} U^{(\leq r')}(k, K s_n) = O_{\mathbb{P}}(s_n^2). \qquad \square$$

Finally, Theorem 2.15 follows from Corollary 7.5, Lemma 7.9, and the fact that the upper bounds on $\mathsf{BP}_{\mathsf{BB}\,\sigma_n}$ in Lemma 7.12 are uniform over $(\sigma_n)_n$:

Proof of Theorem 2.15. Lemma 7.9 allows us to bound $T_{\sigma_n^2\text{-lucky}}$ by $T_{\mathsf{BB}\,K'\sigma_n}$ apart from an event of small probability. Since $T_{\text{size}\,\sigma_n^2} \leq T_{\sigma_n^2\text{-lucky}}$ and $T_{\text{height}\,\sigma_n} \leq T_{\mathsf{BB}\,\sigma_n}$ by construction, the same holds true for the other times from Theorem 2.15. Writing T for any one of these other times, Corollary 7.5 implies that $f_n^{-1}(T) \leq \mathsf{MC}_0 + \varepsilon$ whp. For the matching lower bound, we argue as in the proof of Corollary 7.5. It suffices to observe that $\sigma_n \to \infty$ implies that $|\mathsf{BP}_T| \overset{\mathbb{P}}{\longrightarrow} \infty$, whereas $\left|\mathsf{BP}_{f_n(\mathsf{MC}_0)}\right|$ is bounded by the number of vertices invaded before the edge of PWIT weight MC_0. This latter quantity is tight since it does not depend on n, and therefore $f_n(\mathsf{MC}_0) \geq T$ whp.

The upper bounds on the size and diameter of $\mathsf{BP}_{\mathsf{BB}\,\sigma_n}$ follow from Lemma 7.12. For the lower bound on the diameter, note that $\mathsf{BP}_{\mathsf{BB}\,\sigma_n}$ has diameter at least σ_n by construction. For the lower bound on the size, let $\varepsilon > 0$ be given and let K' be the constant from Lemma 7.9. If $\sigma_n \geq K'$, we can apply Lemma 7.9 with σ_n replaced by σ_n/K' to conclude that $\mathsf{BP}_{\mathsf{BB}\,\sigma_n}$ contains at least $(\sigma_n/K')^2$ vertices with probability at least $1-\varepsilon$. If not, then σ_n is bounded by K' and we can use the trivial bound $|\mathsf{BP}_{\mathsf{BB}\,\sigma_n}| \geq 1$.

Upper bounds on the size and diameter of BP_T follow from the upper bounds on $\mathsf{BP}_{\mathsf{BB}\,\sigma_n}$ and the fact mentioned above that $T \leq T_{\mathsf{BB}\,\sigma_n}$ with probability close to 1, as above.

For the other lower bounds, note first that $\mathsf{BP}_{\mathrm{size}\,\sigma_n^2}$ has size at least σ_n^2 by definition. It suffices to show that $\mathsf{BP}_{\mathrm{size}\,\sigma_n^2}$ contains a vertex of height at least $\delta\sigma_n$ with a probability that can be made arbitrarily close to 1, uniformly over n, by taking δ small enough. More precisely, let $K < \infty$ and $\varepsilon > 0$ be given, and consider any sequence $(\sigma_n)_n$ with $\sigma_n \leq Ks_n$. By Lemma 7.12, there is a constant C such that

$$\mathbb{P}\left(|\mathsf{BP}_{\mathsf{BB}\,\tilde{\sigma}_n}| \geq C\tilde{\sigma}_n^2\right) \leq \varepsilon \tag{7.48}$$

whenever $1 \leq \tilde{\sigma}_n \leq Ks_n$. Now notice that if all vertices of $\mathsf{BP}_{\mathrm{size}\,\sigma_n^2}$ have heights strictly less than $\delta\sigma_n$, then $\mathsf{BP}_{\mathrm{height}\,\delta\sigma_n}$ has size at least σ_n^2. Then, taking $\tilde{\sigma}_n = \delta\sigma_n$ and $\delta = \min\{1/\sqrt{C}, 1\}$ (so that $\tilde{\sigma}_n \leq \sigma_n \leq Ks_n$),

$$\mathbb{P}\left(\mathrm{height}(\mathsf{BP}_{\mathrm{size}\,\sigma_n^2}) < \delta\sigma_n\right) \leq \mathbb{P}(|\mathsf{BP}_{\mathrm{height}\,\tilde{\sigma}_n}| \geq C\tilde{\sigma}_n^2) \leq \mathbb{P}(|\mathsf{BP}_{\mathsf{BB}\,\tilde{\sigma}_n}| \geq C\tilde{\sigma}_n^2) \leq \varepsilon, \tag{7.49}$$

at least if $\tilde{\sigma}_n \geq 1$; on the other hand, if $\tilde{\sigma}_n = \delta\sigma_n < 1$ then there is nothing to prove because $\mathsf{BP}_{\mathrm{size}\,\sigma_n^2}$ must always have diameter at least 1.

The lower bounds for $\mathsf{BP}_{\sigma_n^2\text{-lucky}}$ follow because $\mathsf{BP}_{\mathrm{size}\,\sigma_n^2} \subset \mathsf{BP}_{\sigma_n^2\text{-lucky}}$.

For $\mathsf{BP}_{\mathrm{height}\,\sigma_n}$, note that $\mathsf{BP}_{\mathrm{height}\,\sigma_n}$ has diameter at least σ_n by construction. For the size, we reverse the argument above by noticing that if $|\mathsf{BP}_{\mathrm{height}\,\sigma_n}| < \delta\sigma_n^2$, then $\mathsf{BP}_{\mathrm{size}\,\delta\sigma_n^2}$ has diameter at least σ_n. Let $K < \infty$ and $\varepsilon > 0$ be given, let K' be the constant from Lemma 7.9, and consider any sequence $(\sigma_n)_n$ with $\sigma_n \leq Ks_n$. From Lemma 7.12 we may choose $C' < \infty$ such that

$$\mathbb{P}\left(\mathrm{diameter}(\mathsf{BP}_{\mathsf{BB}\,\hat{\sigma}_n}) \geq C'\hat{\sigma}_n\right) \leq \varepsilon \tag{7.50}$$

whenever $1 \leq \tilde{\sigma}_n \leq Ks_n$. Then, writing $\tilde{\sigma}_n = \sqrt{\delta}\sigma_n$ and $\hat{\sigma}_n = K'\tilde{\sigma}_n = K'\sqrt{\delta}\sigma_n$, and setting $\delta = \min\left\{(C'K')^{-2}, (K')^{-2}, 1\right\}$ (so that $\max\{\tilde{\sigma}_n, \hat{\sigma}_n\} \leq \sigma_n \leq Ks_n$),

$$\begin{aligned}
\mathbb{P}\left(|\mathsf{BP}_{\mathrm{height}\,\sigma_n}| < \delta\sigma_n^2\right) &\leq \mathbb{P}\left(\mathrm{diameter}(\mathsf{BP}_{\mathrm{size}\,\tilde{\sigma}_n^2}) \geq \sigma_n\right) \\
&\leq \mathbb{P}\left(T_{\mathrm{size}\,\tilde{\sigma}_n^2} > T_{\mathsf{BB}\,K'\tilde{\sigma}_n}\right) + \mathbb{P}\left(\mathrm{diameter}(\mathsf{BP}_{\mathsf{BP}\,K'\tilde{\sigma}_n}) \geq \sigma_n\right) \\
&\leq \varepsilon + \mathbb{P}\left(\mathrm{diameter}(\mathsf{BP}_{\mathsf{BP}\,\hat{\sigma}_n}) \geq \sigma_n\right) \leq 2\varepsilon, \tag{7.51}
\end{aligned}$$

since $\sigma_n \geq C'\hat{\sigma}_n$ by construction. $\qquad\square$

7.5 Remaining proofs: Theorems 2.4, 1.1–1.3 and 1.5

In this section, we complete the proof of our main results Theorems 1.1–1.3 and 1.5. We start by proving Theorem 2.4, which applies to the most general edge weights.

Proof of Theorem 2.4. We may use Theorem 2.10 to bound SWT in terms of BP and $\widetilde{\mathsf{BP}}$. In fact, for definiteness, we may assume that BP and SWT are coupled as in (3.4), with $\mathsf{SWT}_t = \pi_M(\widetilde{\mathsf{BP}}_t)$ for all $t \geq 0$.

Consider a time T depending only on the edge weights $X_v, v \in \mathcal{T} \setminus \{\varnothing_1\}$, but not on the marks M_v. For instance, any of the four times from Theorem 2.15 have this property. On $\{v \in \mathsf{BP}_T\}$, we can upper-bound the event $\{v \text{ is thinned}\}$ by the event that v or any

of its $|v|$ ancestors has the same mark as some other vertex $w \in \mathsf{BP}_T$. (This is an upper bound because we ignore the possibility that w is itself thinned or is explored after v.) Since any two vertices have conditional probability $1/n$ of having the same mark, Markov's inequality implies

$$\mathbb{1}_{\{v \in \mathsf{BP}_T\}} \mathbb{P}\left(v \text{ is thinned} \mid T, \mathsf{BP}_T\right) \leq \mathbb{1}_{\{v \in \mathsf{BP}_T\}}(|v|+1)\frac{|\mathsf{BP}_T|}{n} \leq \frac{|\mathsf{BP}_T|\,(\text{diameter}(\mathsf{BP}_T)+1)}{n}. \tag{7.52}$$

Now consider $T = T_{\text{height } \sigma_n}$, $T = T_{\text{size } \sigma_n^2}$ or $T = T_{\text{size } 2\sigma_n^2}$. By Theorem 2.15, the upper bound in (7.52) is $O_{\mathbb{P}}(\sigma_n^3/n)$, which is $o_{\mathbb{P}}(1)$ by the assumption $\sigma_n = o(n^{1/3})$. Hence, if V denotes a vertex of BP_T having maximal height, we conclude that

$$\text{height}(\mathsf{BP}_T) = \text{height}(\widetilde{\mathsf{BP}}_T) \qquad \text{whp.} \tag{7.53}$$

Likewise, summing over v gives

$$\mathbb{E}\left(\left|\mathsf{BP}_T \setminus \widetilde{\mathsf{BP}}_T\right| \,\Big|\, T, \mathsf{BP}_T\right) = o_{\mathbb{P}}(|\mathsf{BP}_T|). \tag{7.54}$$

In particular, we infer that

$$T_{\text{height } \sigma_n}^{(K_n)} = T_{\text{height } \sigma_n} \quad \text{and} \quad T_{\text{size } \sigma_n^2} \leq T_{\text{size } \sigma_n^2}^{(K_n)} \leq T_{\text{size } 2\sigma_n^2} \qquad \text{whp.} \tag{7.55}$$

In view of (7.53)–(7.55), Theorem 2.4 follows from Theorem 2.15. $\qquad\square$

Proof of Theorems 1.1–1.3. These follow immediately from Theorems 2.1, 2.11 and 2.4, respectively, together with Lemmas 5.4–5.5. $\qquad\square$

Proof of Theorem 1.5. The function $f_n(x)$ is given by

$$f_n(x) = (1 - \mathrm{e}^{-x/n})^{s_n} \tag{7.56}$$

(compare (2.2)–(2.3) and (1.21)). We find

$$\frac{x f_n'(x)}{f_n(x)} = s_n \frac{\frac{x}{n}\mathrm{e}^{-x/n}}{1 - \mathrm{e}^{-x/n}} \tag{7.57}$$

and notice that the fraction in the right-hand side of (7.57) is bounded away from 0 and ∞ for x in compact subsets of $[0, \infty)$. As in the proof of Lemma 5.4, this implies Conditions 2.2–2.3, (2.15) and (2.5). Therefore Theorem 1.5 follows from Theorems 2.1, 2.11 and 2.4 and Lemma 5.5. $\qquad\square$

References

[1] Louigi Addario-Berry, *The local weak limit of the minimum spanning tree of the complete graph*, Preprint: arXiv:1301.1667v2 [math.PR], 2013.

[2] Louigi Addario-Berry, Nicolas Broutin, Christina Goldschmidt, and Grégory Miermont, *The scaling limit of the minimum spanning tree of the complete graph*, Ann. Probab. **45** (2017), no. 5, 3075–3144. MR-3706739

[3] Louigi Addario-Berry, Nicolas Broutin, and Bruce Reed, *Critical random graphs and the structure of the minimal spanning tree*, Random Struct. Algor. **35** (2008), no. 3, 323–347. MR-2548517

[4] Louigi Addario-Berry, Simon Griffiths, and Ross J. Kang, *Invasion percolation on the Poisson-weighted infinite tree*, Ann. Appl. Probab. **22** (2012), no. 3, 931–970. MR-2977982

[5] David Aldous, *The continuum random tree. I*, Ann. Probab. **19** (1991), no. 1, 1–28. MR-1085326

[6] David Aldous, *The continuum random tree. III*, Ann. Probab. **21** (1993), no. 1, 248–289. MR-1207226

[7] David Aldous and J. Michael Steele, *The objective method: probabilistic combinatorial optimization and local weak convergence*, Probability on Discrete Structures, Encyclopaedia of Mathematical Sciences, vol. 110, Springer, Berlin, 2004, pp. 1–72. MR-2023650

[8] Omer Angel, Jesse Goodman, Frank den Hollander, and Gordon Slade, *Invasion percolation on regular trees*, Ann. Probab. **36** (2008), no. 2, 420–466. MR-2393988

[9] Itai Benjamini and Oded Schramm, *Recurrence of distributional limits of finite planar graphs*, Electron. J. Probab. **6** (2001), no. 23, 1–13. MR-1873300

[10] Shankar Bhamidi and Remco van der Hofstad, *Weak disorder asymptotics in the stochastic mean-field model of distance*, Ann. Appl. Probab. **22** (2012), no. 1, 29–69. MR-2932542

[11] Shankar Bhamidi and Remco van der Hofstad, *Diameter of the stochastic mean-field model of distance*, Combin. Probab. Comput. **26** (2017), no. 6, 797–825. MR-3714992

[12] Shankar Bhamidi, Remco van der Hofstad, and Gerard Hooghiemstra, *Weak disorder in the stochastic mean-field model of distance II*, Bernoulli **19** (2013), no. 2, 363–386. MR-3037157

[13] Michael Damron and Artëm Sapozhnikov, *Outlets of 2D invasion percolation and multiple-armed incipient infinite clusters*, Probab. Theory Rel. Fields **150** (2011), no. 1-2, 257–294. MR-2800910

[14] Maren Eckhoff, Jesse Goodman, Remco van der Hofstad, and Francesca R. Nardi, *Long paths in first passage percolation on the complete graph II. Global branching dynamics*, to appear in J. Stat. Phys. DOI:10.1007/s10955-020-02585-1.

[15] Maren Eckhoff, Jesse Goodman, Remco van der Hofstad, and Francesca R. Nardi, *Short paths for first passage percolation on the complete graph*, J. Stat. Phys. **151** (2013), 1056–1088. MR-3063496

[16] Daniel Fernholz and Vijaya Ramachandran, *The diameter of sparse random graphs*, Random Structures Algorithms **31** (2007), no. 4, 482–516. MR-2362640

[17] Jesse Goodman, *Exponential growth of ponds in invasion percolation on regular trees*, J. Stat. Phys. **147** (2012), 919–941. MR-2946629

[18] Theodore E. Harris, *The theory of branching processes*, Die Grundlehren der Mathematischen Wissenschaften, Bd. 119, Springer-Verlag, Berlin, 1963. MR-0163361

[19] Remco van der Hofstad, *Random graphs and complex networks. Vol. 1*, Cambridge Series in Statistical and Probabilistic Mathematics, [43], Cambridge University Press, Cambridge, 2017. MR-3617364

[20] Svante Janson, *One, two and three times $\log n/n$ for paths in a complete graph with random weights*, Comb. Probab. Comput. **8** (1999), no. 4, 347–361. MR-1723648

[21] Daniel L. Stein and Charles M. Newman, *Broken ergodicity and the geometry of rugged landscapes*, Phys. Rev. E **51** (1995), no. 6, 5228–5238.

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm