The B_2 index of galled trees

François Bienvenu[1,2], Jean-Jil Duchamps[1],
Michael Fuchs[3], and Tsan-Cheng Yu[3]

[1] *Université de Franche-Comté, CNRS, LmB, F-25000 Besançon, France.*

[2] *Institute for Theoretical Studies, ETH Zürich, 8092 Zürich, Switzerland.*

[3] *Department of Mathematical Sciences, National Chengchi University, Taipei 116, Taiwan.*

August 27, 2024

Abstract

In recent years, there has been an effort to extend the classical notion of phylogenetic balance, originally defined in the context of trees, to networks. One of the most natural ways to do this is with the so-called B_2 index. In this paper, we study the B_2 index for a prominent class of phylogenetic networks: galled trees. We show that the B_2 index of a uniform leaf-labeled galled tree converges in distribution as the network becomes large. We characterize the corresponding limiting distribution, and show that its expected value is 2.707911858984... This is the first time that a balance index has been studied to this level of detail for a random phylogenetic network.

One specificity of this work is that we use two different and independent approaches, each with its advantages: analytic combinatorics, and local limits. The analytic combinatorics approach is more direct, as it relies on standard tools; but it involves slightly more complex calculations. Because it has not previously been used to study such questions, the local limit approach requires developing an extensive framework beforehand; however, this framework is interesting in itself and can be used to tackle other similar problems.

Contents

1 Introduction **2**
1.1 Biological context and main result 2
1.2 Setting and notation 3

2 Analytic combinatorics approach **5**

3 Local limit approach **10**
3.1 General principle 10
3.2 The B_2 index of infinite phylogenetic networks 11
3.3 Blowups of Galton–Watson trees 14
3.4 Application to leaf-labeled galled trees 18

References **20**

Appendices **23**
A.1 The limit law via the method of moments 23
A.2 The boundary of a phylogenetic network 28
A.3 Properties of the entropy 31
A.4 Properties of the B_2 index 34
A.5 Total variation bound for the convergence of Galton–Watson trees . 37
A.6 Blowups of Galton–Watson trees: proofs 40

1 Introduction

1.1 Biological context and main result

Indices of phylogenetic balance – *balance indices* for short – are summary statistics that quantify the idea that some trees have more symmetries than others. They play a central role in theoretical evolutionary biology, and are also widely used in practical applications. There is a great diversity of balance indices, and their mathematical properties are for the most part well-understood; see [10] for a comprehensive survey.

However, with the growing recognition of the importance of reticulate evolution, phylogenetic networks are increasingly used to describe phylogenies. Extending the definition and study of balance indices to networks is thus becoming an important topic in mathematical phylogenetics, and very recently some authors have started studying extensions of existing balance indices in networks [29] and designing new balance indices specifically with networks in mind [19].

In that context, one balance index known as the B_2 index stands out. Introduced in the context of trees in [24], this balance index had gone largely unnoticed. However, as recently pointed out in [4], where it was studied extensively, its formal definition and its interpretation as a balance index are unchanged in the context of networks; we come back to this in Section 1.2 below.

Although it has been studied for various models of random trees, the distribution of the B_2 index has not been studied for random phylogenetic networks. In this paper, we do so for a natural and well-studied model: uniform leaf-labeled galled trees (note that there is some variation regarding the term *galled tree* – see [8] for a detailed discussion – and that our use of the term follows the prevalent usage [14] and refers to what are also known as *level-1 networks*).

We show that the B_2 index of a galled tree G_n sampled uniformly at random among galled trees with n labeled leaves converges in distribution as n goes to infinity, and we give a characterization of the limiting distribution. We also show the convergence of all p-th moments and that, in particular, $\mathbb{E}(B_2(G_n))$ converges to a constant $c = 2.707911858984\ldots$

This is the first result of this type for the balance index of a random phylogenetic network: the only comparable result in the current literature is that the expected value of an extension of the Sackin index of a uniform simplex network with n labeled leaves is asymptotically $\Theta(n^{7/4})$, as proved in [29]. However, beyond the fact that this result is less precise than ours, it is not clear that the extension of the Sackin index considered in that paper is a measure of phylogenetic balance. More results concerning the distribution of balances indices in phylogenetic networks will be published in a forthcoming paper by one of the authors [12]. In this paper, the limiting distribution (after scaling) of various extensions of the Sackin index is studied for several models of phylogenetic networks – including the galled trees considered here – and is shown to be the Airy distribution.

A notable feature of this paper is that we use two very different – and completely independent – approaches to study the asymptotic behavior of $B_2(G_n)$:

- a combinatorial approach, based on analytic combinatorics;
- a probabilistic approach, based on modern tools from the study of branching processes: local limits of conditioned Galton–Watson trees.

The point of presenting these two approaches in the same paper, even though it means proving the main results twice, is that each approach has its advantages and disadvantages. The combinatorial approach is more direct, because it relies on a well-established framework (namely, the method of *singularity analysis*; see [11]). However, it involves calculations that are very specific to the problem at hand; and while it is possible to characterize the limiting distribution of $B_2(G_n)$ through its moments, this requires increasingly strenuous calculations as the moments get higher. The probabilistic approach, on the other hand, requires establishing a substantial number of technical preliminaries. This is because, although local limits have become a standard tool in probability theory, there are difficulties regarding the definition of B_2 for infinite phylogenetic networks and its continuity for the local topology. This preliminary work is thus needed to justify the rigorousness of the approach; but the numerical calculations themselves are then fairly simple, and immediately yield a simple recursive characterization of the limiting distribution of $B_2(G_n)$. Moreover, now that the technicalities have been worked out, it is straightforward to apply this approach to any other random phylogenetic network with a suitable local limit.

In the rest of this section, we formally define the B_2 index and the class of galled trees. We then split the paper into two independent sections: Section 2 details the combinatorial approach, and Section 3 the probabilistic one. In order to emphasize the practicality of the latter as a tool to perform concrete computations, only the main ideas of the framework and its application to galled trees are presented in the main text; its technical justification is detailed in the Appendix.

1.2 Setting and notation

Let us start by specifying what we mean by *phylogenetic network*. In order to prepare the ground for the local limit approach, we include infinite graphs in our definition.

Recall that a DAG is a directed graph with no directed cycles. It is:

- *countable* if its vertex set is finite or countably infinite;
- *locally finite* if every vertex has finite in-degree and out-degree;
- *rooted* if there is a unique vertex with in-degree 0, called the *root*, and every vertex v is reachable from the root (that is, there exists a directed path going from the root to v).

Definition 1.1. A *phylogenetic network* is a countable, locally finite rooted DAG. Its vertices with out-degree 0, if any, are called the *leaves*. ◇

Next, we recall the definition of the B_2 index of a *finite* phylogenetic network, as given in [4]. Extending this definition to infinite phylogenetic networks will be one of the main challenges for the local limit approach.

Let G be a finite phylogenetic network, and let $X = (X_t)_{t \geqslant 0}$ be the simple random walk on G, started from the root and constrained to follow the direction of the edges: at each step, one of the outgoing edges of the current vertex is chosen uniformly at random, until a leaf is reached. In the rest of this document, we will refer to X simply as *the directed random walk on G*. For each vertex $v \in G$, let p_v denote the probability that X visits v. In a finite phylogenetic network, $(p_\ell)_{\ell \in \mathcal{L}}$ is a probability distribution on $\mathcal{L}$, the leaf set of G.

Definition 1.2. Let G be a finite phylogenetic network. The B_2 *index of* G is the Shannon entropy of the probability distribution $(p_\ell)_{\ell \in \mathcal{L}}$ induced on the leaves of G by the directed random walk – that is,

$$B_2(G) \;=\; -\sum_{\ell \in \mathcal{L}} p_\ell \log_2 p_\ell \,. \qquad \diamond$$

We only recall, without proof, an elementary but fundamental property of the B_2 index that will be of constant use in this paper. We refer the reader to [4] for a more detailed presentation.

Proposition 1.3 (grafting property)**.** *Let G_1 and G_2 be two finite phylogenetic networks, and let G be the phylogenetic network obtained by identifying a leaf $\ell \in G_1$ with the root of G_2. Then,*

$$B_2(G) \;=\; B_2(G_1) \;+\; p_\ell \, B_2(G_2) \,,$$

where p_ℓ denotes the probability that the directed random walk on G_1 ends in ℓ.

Let us close this section by formally defining the class of galled trees that we consider. Since in this paper we work in the context of DAGs, we use the term *connectivity* to refer to the notion of *weak connectivity* (that is, connectivity of the underlying undirected graph): there is no risk of confusion with the notion of strong connectivity, as this notion makes little sense for DAGs (whose strongly connected components are reduced to single vertices).

Recall that v is a *cut-vertex* of G if removing v increases the number of connected components of G. A graph is said to be *biconnected* if it has no cut-vertices, and a *biconnected component* is a maximal biconnected subgraph. A biconnected component consisting of a single vertex is said to be trivial.

Finally, a phylogenetic network G is *binary* if it is either reduced to a single vertex, or if the root has out-degree 2; the leaves have in-degree 1; and every other vertex has either in-degree 1 and out-degree 2 (*tree-vertices*) or in-degree 2 and out-degree 1 (*reticulations*).

Definition 1.4. A *galled tree*, also known as a *level-1 network*, is a binary phylogenetic network with no multi-edge where each biconnected component has at most one reticulation. ◇

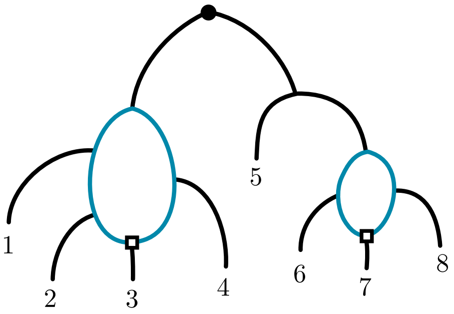

Figure 1: An example of a galled network with $n = 8$ labeled leaves and two galls. The root is the black circle on top, and each point where three lines meet is an internal vertex. The direction of the edges is not indicated: they are always pointing downwards. The two galls are highlighted in blue, and the corresponding reticulations are denoted by white squares. The probability that the directed random walk ends in leaf 1 is $p_1 = 1/8$; the probability that it ends in leaf 3 is $p_3 = 1/16 + 1/8$.

The non-trivial biconnected components of a galled tree are known as its *galls*. They correspond to undirected cycles made of two directed paths connecting a vertex, known as the *root of the gall*, to a reticulation. Note that the galls of a galled tree are, by definition, vertex-disjoint.

Our definitions so far make no mention of labelings: they apply to unlabeled, vertex-labeled and leaf-labeled networks alike. However, the labeling plays an important role in the definition of our model of random phylogenetic network.

Let $\mathscr{G}_n$ denote the combinatorial class of galled trees with n leaves labeled with the integers from 1 to n. Note that $\mathscr{G}_n$ is finite: the sequence $(\operatorname{Card} \mathscr{G}_n)_{n \geqslant 1}$ is registered as A328122 in the On-Line Encyclopedia of Integer Sequences [1], and has been studied extensively in [7], where an explicit formula was obtained using analytic combinatorics. Recently, Stufler [26] studied $\mathscr{G}_n$ (and, more generally, level-k networks) using branching processes; his method is the starting point of our probabilistic approach, and we will come back to it later.

Since $\mathscr{G}_n$ is finite, we can endow it with the uniform distribution: we write $G_n \sim \operatorname{Unif}(\mathscr{G}_n)$ to denote a phylogenetic network G_n sampled uniformly at random from $\mathscr{G}_n$, and we refer to G_n as a *uniform galled tree with n labeled leaves*. Our main object of study in the rest of this document is the asymptotic behavior of the random variable $B_2(G_n)$.

2 Analytic combinatorics approach

In this section, we derive the asymptotics of moments of $B_2(G_n)$. To do so, we use a generating function approach that is based on the recursive decomposition of galled trees according to whether the root is in a gall or not; see Figure 2. This decomposition yields explicit expressions for the generating functions of moments, which can then be analysed with the method of singularity analysis from Analytic Combinatorics; see Chapter VI in [11]. The latter is based on so-called transfer theorems which say, in a nutshell, that if a generating function $f(z)$ is analytic in a suitable large domain Δ with a singularity $\xi \in \mathbb{R}^+$ at its boundary and $f(z) \sim c(1 - z/\xi)^{-\alpha}$, as $z \to \xi$ in Δ, then $[z^n]f(z) \sim c n^{\alpha-1} \xi^{-n} / \Gamma(\alpha)$, as $n \to \infty$,

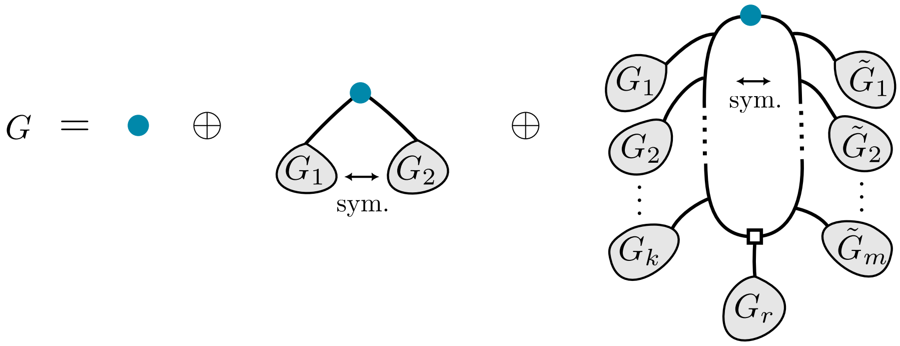

Figure 2: The recursive decomposition of a galled tree G. G is either a leaf (first case after the equality sign) or a root which is not contained in a gall, with two galled trees G_1 and G_2 attached to it (second case); or the root itself is in a gall (third case; both sides of the gall are not allowed to be empty at the same time, i.e. $k+m \geqslant 1$). Note that in the second and third cases, there is a symmetry: it does not matter which side of the gall is drawn to the left *vs* to the right.

where $[z^n]f(z)$ denotes the n-th Taylor coefficient of $f(z)$ at $z=0$. More precisely, we use domains Δ of the form:

$$\Delta := \Delta(r,\varphi_0) = \{z \;:\; |z| \leqslant r,\; z \neq \xi,\; |\arg(z-\xi)| \geqslant \varphi_0\}$$

with $r > |\xi|$ and $0 < \varphi_0 < \pi/2$. A function which is analytic in Δ (for some r and φ_0) is subsequently called Δ-*analytic at* ξ.

We first need to recall some results from [7]. Set $\mathscr{G} := \bigcup_{n\geqslant 0}\mathscr{G}_n$ which is the set of all galled trees. Moreover, set

$$\mathrm{GT}(z) := \sum_{G\in\mathscr{G}} \frac{z^{|G|}}{|G|!},$$

which is the (exponential) generating function of the number of galled trees G with $|G| =: n$ leaves. Then, from [7],

$$\mathrm{GT}(z) = -\frac{\sqrt{2}\sqrt{-4\left(\sqrt{1-8z}-2\right)z+9\sqrt{1-8z}-1}}{4(1-8z)^{1/4}} - \frac{1}{4}\sqrt{1-8z}+\frac{5}{4}. \tag{1}$$

Thus, for a Δ-domain at 1/8, $\mathrm{GT}(z)$ is Δ-analytic with an expansion at $z = 1/8$ of the form:

$$\mathrm{GT}(z) = \rho - \tau\sqrt{1-8z} + \mathcal{O}(1-8z), \qquad (z \to 1/8). \tag{2}$$

The above limit is inside the Δ-domain and

$$\rho = \frac{5}{4} - \frac{\sqrt{17}}{4} \approx 0.219 \qquad \text{and} \qquad \tau = \frac{1}{4} - \frac{\sqrt{17}}{68} \approx 0.189.$$

From this, by the transfer theorems,

$$|\mathscr{G}_n| = n![z^n]\mathrm{GT}(z) \sim -\tau n![z^n]\sqrt{1-8z} \sim \frac{\sqrt{2}\tau}{2}\left(\frac{8}{e}\right)^n n^{n-1};$$

compare with Proposition 5.3 in [7].

Formula (1) was also derived with the decomposition mentioned above. (Actually, the authors in [7] divided the third case in Figure 2 into two subcases, as this allowed an easier translation into generating function.) We now extend the analysis to the mean.

Set:

$$A(z) := \sum_{G\in\mathcal{G}} B_2(G)\frac{z^{|G|}}{|G|!},$$

where $B_2(G)$ is the B_2 index of the galled tree G whose definition we recall:

$$B_2(G) = -\sum_{\ell\in\mathcal{L}(G)} p_\ell \log_2 p_\ell,$$

where $\mathcal{L}(G)$ denotes the leaf set of G and p_ℓ is the probability of reaching ℓ.

First, consider the case where the root of G is not in a gall. Then, it has two galled subtrees G_1, G_2, and we have

$$B_2(G) = \frac{1}{2}\left(B_2(G_1) + B_2(G_2)\right) + 1 \tag{3}$$

as directly follows from the grafting property.

Next, consider the case where the root of G is inside a gall. Assume that $G_1, \ldots, G_k$ and $\tilde{G}_1, \ldots, \tilde{G}_m$ are the galled subtrees attached to the left and right of the cycle (in any order) and G_r the galled subtree below the reticulation vertex at the bottom of the gall. Then, for $k, m \geqslant 0$ with $k + m \geqslant 1$, we have

$$\begin{aligned}
B_2(G) &= -\sum_{s=1}^{k}\sum_{\ell\in\mathcal{L}(G_s)} \frac{p_\ell}{2^{s+1}}\log_2\left(\frac{p_\ell}{2^{s+1}}\right) - \sum_{t=1}^{m}\sum_{\ell\in\mathcal{L}(\tilde{G}_t)} \frac{p_\ell}{2^{t+1}}\log_2\left(\frac{p_\ell}{2^{t+1}}\right)\\
&\qquad - \sum_{\ell\in\mathcal{L}(G_r)}\left(\frac{1}{2^{k+1}} + \frac{1}{2^{m+1}}\right) p_\ell \log_2\left(\left(\frac{1}{2^{k+1}} + \frac{1}{2^{m+1}}\right)p_\ell\right)\\
&= \sum_{s=1}^{k}\frac{B_2(G_s)}{2^{s+1}} + \sum_{t=1}^{m}\frac{B_2(\tilde{G}_t)}{2^{t+1}} + \left(\frac{1}{2^{k+1}} + \frac{1}{2^{m+1}}\right)B_2(G_r)\\
&\qquad + \sum_{s=1}^{k}\frac{s+1}{2^{s+1}} + \sum_{t=1}^{m}\frac{t+1}{2^{t+1}} - \left(\frac{1}{2^{k+1}} + \frac{1}{2^{m+1}}\right)\log_2\left(\frac{1}{2^{k+1}} + \frac{1}{2^{m+1}}\right)\\
&= \sum_{s=1}^{k}\frac{B_2(G_s)}{2^{s+1}} + \sum_{t=1}^{m}\frac{B_2(\tilde{G}_t)}{2^{t+1}} + \left(\frac{1}{2^{k+1}} + \frac{1}{2^{m+1}}\right)B_2(G_r)\\
&\qquad + 3 - \frac{k+3}{2^{k+1}} - \frac{m+3}{2^{m+1}} - \left(\frac{1}{2^{k+1}} + \frac{1}{2^{m+1}}\right)\log_2\left(\frac{1}{2^{k+1}} + \frac{1}{2^{m+1}}\right).
\end{aligned}\tag{4}$$

We now use the recursive decomposition, which gives:

$$\begin{aligned}
A(z) &= \frac{1}{2}\sum_{G_1,G_2} B_2(G)\frac{z^{|G_1|+|G_2|}}{|G_1|!|G_2|!}\\
&\qquad + \frac{1}{2}\sum_{\substack{k,m\geqslant 0\\ k+m\geqslant 1}}\ \sum_{\substack{G_1,\ldots,G_k\\ \tilde{G}_1,\ldots,\tilde{G}_m\\ G_r}} B_2(G)\frac{z^{|G_1|+\cdots+|G_k|+|\tilde{G}_1|+\cdots+|\tilde{G}_m|+|G_r|}}{|G_1|!\cdots|G_k|!|\tilde{G}_1|!\cdots|\tilde{G}_m|!|G_r|!},
\end{aligned}\tag{5}$$

where inside the first and second sum, we have to replace $B_2(G)$ by (3) and (4), respectively. This gives, for the first sum,

$$\sum_{G_1,G_2} B_2(G)\frac{z^{|G_1|+|G_2|}}{|G_1|!|G_2|!} = A(z)\mathrm{GT}(z) + \mathrm{GT}(z)^2.$$

For the second sum, we have

$$\begin{aligned}
&\sum_{\substack{k,m\geqslant 0\\ k+m\geqslant 1}} \sum_{\substack{G_1,\dots,G_k\\ \tilde{G}_1,\dots,\tilde{G}_m\\ G_r}} B_2(G)\frac{z^{|G_1|+\cdots+|G_k|+|\tilde{G}_1|+\cdots+|\tilde{G}_m|+|G_r|}}{|G_1|!\cdots|G_k|!|\tilde{G}_1|!\cdots|\tilde{G}_m|!|G_r|!}\\
&\quad = A(z)\sum_{\substack{k,m\geqslant 0\\ k+m\geqslant 1}} \left(\sum_{s=1}^{k}\frac{1}{2^{s+1}} + \sum_{t=1}^{m}\frac{1}{2^{t+1}} + \left(\frac{1}{2^{k+1}}+\frac{1}{2^{m+1}}\right)\right)\mathrm{GT}(z)^{k+m}\\
&\qquad + \sum_{\substack{k,m\geqslant 0\\ k+m\geqslant 1}} \left(3 - \frac{k+3}{2^{k+1}} - \frac{m+3}{2^{m+1}}\right)\mathrm{GT}(z)^{k+m+1}\\
&\qquad - \sum_{\substack{k,m\geqslant 0\\ k+m\geqslant 1}} \left(\frac{1}{2^{k+1}}+\frac{1}{2^{m+1}}\right)\log_2\left(\frac{1}{2^{k+1}}+\frac{1}{2^{m+1}}\right)\mathrm{GT}(z)^{k+m+1}\\
&\quad = A(z)\sum_{\substack{k,m\geqslant 0\\ k+m\geqslant 1}} \mathrm{GT}(z)^{k+m} + \sum_{k,m\geqslant 0}\left(3 - \frac{k+3}{2^{k+1}} - \frac{m+3}{2^{m+1}}\right)\mathrm{GT}(z)^{k+m+1}\\
&\qquad - \sum_{k,m\geqslant 0} \left(\frac{1}{2^{k+1}}+\frac{1}{2^{m+1}}\right)\log_2\left(\frac{1}{2^{k+1}}+\frac{1}{2^{m+1}}\right)\mathrm{GT}(z)^{k+m+1}\\
&\quad = A(z)\left(\frac{1}{(\mathrm{GT}(z)-1)^2} - 1\right) - \frac{\mathrm{GT}(z)^2(\mathrm{GT}(z)-4)}{(\mathrm{GT}(z)-1)^2(\mathrm{GT}(z)-2)^2} - h(\mathrm{GT}(z)),
\end{aligned}$$

where

$$h(\omega) = \sum_{k,m\geqslant 0}\left(\frac{1}{2^{k+1}}+\frac{1}{2^{m+1}}\right)\log_2\left(\frac{1}{2^{k+1}}+\frac{1}{2^{m+1}}\right)\omega^{k+m+1}.$$

Plugging everything into (5) and solving for $A(z)$ gives

$$A(z) = \frac{f(\mathrm{GT}(z))}{g(\mathrm{GT}(z))},$$

where

$$g(\omega) = 1 - \frac{\omega}{2} + \frac{\omega(\omega-2)}{2(\omega-1)^2}$$

and

$$f(\omega) = \frac{\omega^2}{2} - \frac{\omega^2(\omega-4)}{2(\omega-1)^2(\omega-2)^2} - \frac{h(\omega)}{2}.$$

We next show that $A(z)$ is Δ-analytic (with a Δ-domain at $1/8$) and derive its expansion as $z \to 1/8$. First note, that $f(\omega)/g(\omega)$ is analytic in $|\omega| < 1$. Expanding at $\rho \approx 0.219$ gives

$$\frac{f(\omega)}{g(\omega)} = \frac{f(\rho)}{g(\rho)} + \frac{f'(\rho)g(\rho) - f(\rho)g'(\rho)}{g(\rho)^2}(\omega-\rho) + \mathcal{O}((\omega-\rho)^2).$$

By plugging into this (2),

$$A(z) = \frac{f(\rho)}{g(\rho)} + \frac{f(\rho)g'(\rho) - f'(\rho)g(\rho)}{g(\rho)^2}\tau\sqrt{1-8z} + \mathcal{O}((1-8z)), \qquad (z \to 1/8).$$

Thus, by the transfer theorems,

$$[z^n]A(z) \sim \frac{f(\rho)g'(\rho) - f'(\rho)g(\rho)}{g(\rho)^2}\tau[z^n]\sqrt{1-8z}$$

which implies for the mean of the B_2 index of a random galled tree $G_n \sim \mathrm{Unif}(\mathcal{G}_n)$:

$$\mathbb{E}(B_2(G_n)) = \frac{[z^n]A(z)}{[z^n]\mathrm{GT}(z)} \sim \frac{f'(\rho)g(\rho) - f(\rho)g'(\rho)}{g(\rho)^2} = 2.707911858984\ldots,$$

where the numerical evaluation was done with Maple. We summarize this in our first result of this section.

Theorem 2.1. *The mean value of the B_2 index of a uniform galled tree with n labeled leaves converges to* $2.707911858984\ldots$.

The above computations can be generalized to higher moments. More precisely, define

$$A^{[\ell]}(z) := \sum_{G \in \mathcal{G}} B_2(G)^{\ell} \frac{z^{|G|}}{|G|!}$$

so that $A^{[0]}(z) = \mathrm{GT}(z)$ and $A^{[1]}(z) = A(z)$. Then, by raising (3) and (4) to the ℓ-th power, plugging them into (5), and expanding, we obtain that

$$A^{[\ell]}(z) = \frac{f_\ell(z)}{g_\ell(\mathrm{GT}(z))},$$

where

$$\begin{aligned} g_\ell(\omega) := 1 - \frac{\omega}{2^\ell} &- \sum_{k \geqslant 1} \left(\sum_{s=1}^{k} \frac{1}{2^{(s+1)\ell}} + \left(\frac{1}{2} + \frac{1}{2^{k+1}} \right)^{\ell} \right) \omega^k \\ &- \frac{1}{2} \sum_{k,m \geqslant 1} \left(\sum_{s=1}^{k} \frac{1}{2^{(s+1)\ell}} + \sum_{t=1}^{m} \frac{1}{2^{(t+1)\ell}} + \left(\frac{1}{2^{k+1}} + \frac{1}{2^{m+1}} \right)^{\ell} \right) \omega^{k+m} \end{aligned}$$

and

$$\begin{aligned} f_\ell(z) := &\frac{1}{2} \sum_{\substack{\ell_1+\ell_2+\ell_3=\ell \\ \ell_1,\ell_2<\ell}} \binom{\ell}{\ell_1, \ell_2, \ell_3} \frac{1}{2^{\ell_1+\ell_2}} A^{[\ell_1]}(z) A^{[\ell_2]}(z) \\ &+ \frac{1}{2} \sum_{\substack{k,m \geqslant 0 \\ k+m \geqslant 1}} \sum_{\substack{\ell_1+\cdots+\ell_{k+m+2}=\ell \\ \ell_j<\ell, 1 \leqslant j \leqslant k+m+1}} \binom{\ell}{\ell_1, \ldots, \ell_{k+m+2}} \prod_{s=1}^{k} \frac{1}{2^{(s+1)\ell_s}} \prod_{t=1}^{m} \frac{1}{2^{(t+1)\ell_{k+t}}} \\ &\qquad \times \left(\frac{1}{2^{k+1}} + \frac{1}{2^{m+1}} \right)^{\ell_{k+m+1}} \mu_{k,m}^{\ell_{k+m+2}} \prod_{s=1}^{k} A^{[\ell_s]}(z) \prod_{t=1}^{m+1} A^{[\ell_{k+t}]}(z), \end{aligned} \tag{6}$$

where

$$\mu_{k,m} := 3 - \frac{k+3}{2^{k+1}} - \frac{m+3}{2^{m+1}} - \left(\frac{1}{2^{k+1}} + \frac{1}{2^{m+1}} \right) \log_2 \left(\frac{1}{2^{k+1}} + \frac{1}{2^{m+1}} \right). \tag{7}$$

Now, define two sequences c_ℓ and d_ℓ by

$$A^{[\ell]}(z) = c_\ell - d_\ell \tau \sqrt{1-8z} + \cdots$$

Then, we have $c_0 = \rho, d_0 = 1$, and for $\ell \geqslant 1$,

$$c_\ell = \frac{f_\ell(A^{[j]}(z) \leftrightarrow c_j)}{g_\ell(\rho)}$$

and

$$d_\ell = \frac{f'_\ell((A^{[j]})'(z) \leftrightarrow d_j, A^{[j]}(z) \leftrightarrow c_j)}{g_\ell(\rho)} - c_\ell \frac{g'_\ell(\rho)}{g_\ell(\rho)}, \tag{8}$$

where $f_\ell(A^{[j]}(z) \leftrightarrow c_j)$ means that in $f_\ell(z)$ we replace $A^{[j]}(z)$ with $j < \ell$ by c_j and similar for the numerator of the first fraction on the right-hand side of (8).

Finally, by applying singularity analysis as above, we obtain that

$$\mathbb{E}(B_2(G_n)^\ell) \to d_\ell. \tag{9}$$

The above recurrence for d_ℓ can be, e.g., used to compute $d_2 = 7.965561677362\ldots$ (again, e.g., with Maple). Thus, we have the following result for the variance.

Theorem 2.2. *The variance of the B_2 index of a uniform galled tree with n labeled leaves converges to* $0.632774946963\ldots$.

Also, from (9) and the method of moments, we can identify the limit law.

Theorem 2.3. *There exists a random variable B with $\mathbb{E}(B^\ell) = d_\ell$ whose distribution is the limit distribution of the B_2 index of a uniform galled tree with n labeled leaves, i.e.,*

$$B_2(G_n) \xrightarrow{d} B,$$

where the convergence also holds for all moments.

Remark 2.4. An explicit construction of the random variable B from Theorem 2.3 is given in Section 3, where it is obtained as the B_2 index of the infinite phylogenetic network G_* corresponding to the local limit of uniform leaf-labeled galled trees. ◇

Proof. According to classical results from probability theory (see [5, Section 30]), we only have to show that the exponential generating function of $\sum_{\ell \geqslant 0} d_\ell z^\ell/\ell!$ has a non-zero radius of convergence. This follows from an estimate of the form

$$d_\ell \leqslant K^\ell \ell!, \tag{10}$$

where $K > 0$ is a suitable constant. Such an estimate is derived from the recurrence for d_ℓ and induction; see Appendix A.1. □

3 Local limit approach

3.1 General principle

Local limits make it possible to capture certain aspects of the structure of large graphs. Typically, one considers a sequence (G_n) of finite graphs of increasing size. The idea is then to define a (potentially infinite) rooted graph $G_\infty = \lim_n G_n$ whose structure around the root tends to match the local structure of G_n "as seen from a focal vertex" when n goes to infinity. This focal vertex can be fixed or random, depending on the application.

Let us be more specific and briefly recall the definition of local convergence. First, we need to introduce a notion of *restriction* of a phylogenetic network.

Definition 3.1. Let G be a phylogenetic network and let v be a vertex of G. The *height* of v, which we denote by $h(v)$, is the number of edges of a shortest path going from the root of G to v. The *restriction of height k* of G, which we denote by $[G]_k$, is the subgraph of G induced by its vertices of height at most k. ◇

Note that, because phylogenetic networks are locally finite, $[G]_k$ is finite for every phylogenetic network G and every $k \geqslant 0$. Therefore, it is natural to endow the set of finite phylogenetic networks with the discrete topology and thus say that $[G_n]_k \to [G]_k$ when there exists N such that for all $n \geqslant N$, $[G_n]_k = [G]_k$.

Definition 3.2. A sequence (G_n) of phylogenetic networks *converges locally* to the phylogenetic network G if, for all fixed $k \geqslant 0$, as n goes to infinity, $[G_n]_k$ converges to $[G]_k$ in the discrete topology. ◇

The notion of local convergence has now become a standard tool in probability theory, and we refer the reader to [27, Chapter 2] for a detailed introduction. Its usefulness comes in great part from the fact that the local limit G_∞ of a sequence (G_n) of graphs is often more tractable than the finite graphs G_n. In particular, regions of G_n that are only "asymptotically independent" can become truly independent in G_∞. As a result, given a functional f of interest, it is sometimes much easier to compute $f(G_\infty)$ than $f(G_n)$ and this can yield a simple way to compute $\lim_n f(G_n)$. However, for this one must:

- Identify the limit G_∞ as a tractable graph.
- Ensure that f is well-defined for infinite graphs, and that it is continuous for the local topology (or at least along the sequence of interest).

In the case of galled trees, the local limit has been identified in [26] and is indeed highly tractable – see Section 3.3. However, the extension of the B_2 index to infinite phylogenetic networks and its continuity are challenging.

3.2 The B_2 index of infinite phylogenetic networks

The Definition 1.2 of the B_2 index of a finite phylogenetic network as the Shannon entropy of the probability distribution induced on its leaves by the directed random walk X does not immediately extend to infinite networks. This is because the random walk can escape to infinity, without ever ending in a leaf (in fact, note that an infinite phylogenetic network may not even have leaves).

In [4, Definition 1.4], it was claimed that a simple way to circumvent this problem is to define the B_2 index of an infinite phylogenetic network G as $\lim_k B_2([G]_k)$. However, while this works for trees (Proposition A.20 in Appendix A.4), this is not the case for general phylogenetic networks, because the sequence $B_2([G]_k)$ may not converge, as the simple example given in Figure 3 shows.

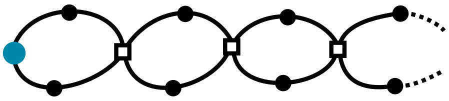

Figure 3: A phylogenetic network G illustrating why the Definition 1.4 from [4] may fail. The root is the vertex highlighted in blue on the right, and the white squares denote reticulations. Here, $B_2([G]_k) = \mathbb{1}_{\{k \text{ is odd}\}}$ does not converge.

To get a general definition, we define a suitable notion of "boundary" of DAGs on which the random walk X induces a probability distribution. In the finite case, this boundary is simply the leaf set; but in the infinite case it consists of the leaf set and of all "directions" in which X can escape. While there already exist several notions of boundaries to quantify the asymptotic behavior of random walks, such as the Martin boundary [17, Chapter 10], we are unaware of a notion of boundary that matches ours. Moreover, here we give an explicit construction that is useful to understand the properties of the B_2 index.

We now give the formal definition of the B_2 index of an arbitrary phylogenetic network, and list some of its most important properties. First, we recall the notion of *ends* of an infinite graph. Note that here we are working with directed graphs, and therefore that the notion of end that we use differs from the more common notion of end of an undirected graph.

A *ray* is a one-way infinite directed path $v_0 \to v_1 \to \cdots$. Two rays r and r' are said to be *co-directional*, which we denote by $r \mathrel{\#\!\!\#} r'$, if there exists a ray r'', which could be one of r or r', that intersects r and r' an infinite number of times. It is readily checked that $\mathrel{\#\!\!\#}$ is an equivalence relation.

Definition 3.3. Let G be a DAG with at least one ray. The equivalence classes of the co-directional relation $\mathrel{\#\!\!\#}$ are called the *ends* of G. ◇

An infinite DAG that is not locally finite may or may not have ends. However, an infinite phylogenetic network, being locally finite, must have at least one end (since it has at least one ray). Moreover, for any $\kappa \in \mathbb{N} \cup \{\aleph_0, 2^{\aleph_0}\}$, there exist phylogenetic networks with κ ends, illustrated in Figure 4.

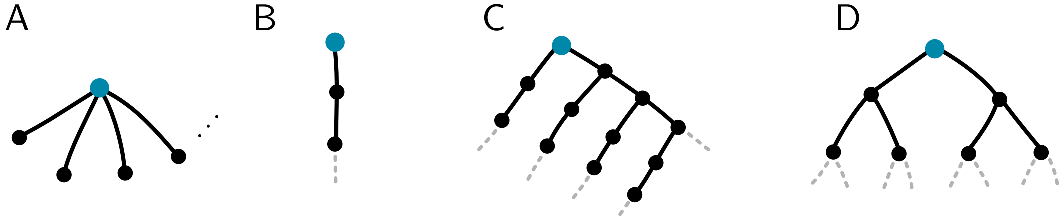

Figure 4: Examples of infinite DAGs with various number of ends. The root is highlighted in blue, and the direction of the edges is not indicated, as they are always oriented away from the root. A: the infinite star graph, an infinite DAG with no ends. This DAG is not locally finite, and therefore is not a phylogenetic network. B: the semi-infinite path, a DAG with one end. C: the DAG obtained by grafting infinite paths to the leaves of the infinite caterpillar. This DAG has $\aleph_0 = \mathrm{Card}(\mathbb{N})$ ends. D: the infinite complete binary tree, a DAG with $2^{\aleph_0} = \mathrm{Card}(\mathbb{R})$ ends.

Definition 3.4. Let G be phylogenetic network with leaf set $\mathcal{L}$ and end set $\mathcal{E}$. The *boundary* of G is the set $\partial G = \mathcal{L} \cup \mathcal{E}$. ◇

In Appendix A.2, we prove that the limit of the random walk $(X_t)_{t\geqslant 0}$ is a well-defined random variable X_∞ taking values in ∂G. We denote its distribution by μ. Using the usual definition of the Shannon entropy H of an arbitrary probability distribution (see Appendix A.3), this makes it possible to extend Definition 1.2 to infinite phylogenetic network.

Definition 3.5. Let G be a phylogenetic network. The B_2 *index of* G is the Shannon entropy of the probability distribution μ induced on the boundary of G by the directed random walk: $B_2(G) = H(\mu)$. $\diamond$

The properties of the B_2 index of infinite phylogenetic networks are essentially the same as in the finite case, except that B_2 can now equal $+\infty$. In particular, the grafting property still holds for infinite phylogenetic networks. This is proved in Appendix A.4, where other basic properties of the B_2 index of general phylogenetic networks are provided.

There is, however, one major difficulty with our approach: the B_2 index is not continuous for the local topology. In fact, B_2 is nowhere continuous on the set of infinite phylogenetic networks – in the sense that for every infinite phylogenetic network G there exists a sequence (G_n) such that $G_n \to G$ for the local topology but $B_2(G_n) \not\to B_2(G)$, see Figure 5. This is because the B_2 index reflects the structure of the boundary of a network, and parts of the boundary can be located infinitely far away from the root; whereas functions that are continuous for the local topology must, by definition, depend on the local structure of the network around the root.

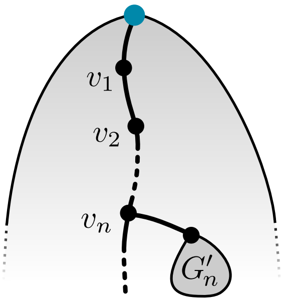

Figure 5: Let G be an infinite phylogenetic network, and let $v_1 \to v_2 \to \cdots$ be a ray of G (being infinite and locally finite, G always has at least one ray). Pick a sequence (G'_n) of phylogenetic networks, and let G_n be the phylogenetic network obtained by grafting G'_n on v_n. Thus, $G_n \to G$. However, by the grafting property, $B_2(G_n) = B_2(G) + p_{v_n} B_2(G'_n)$ can be arbitrarily large.

Nevertheless, it is still possible to use local limits to study the B_2 index: indeed, B_2 can be continuous along sequences (G_n) of interest – in fact, we expect that this should be the case for many biologically relevant models of phylogenetic networks. In Appendix A.4, we provide general results to study the continuity of B_2. For instance, Proposition A.17 states that if the local limit G of a sequence (G_n) is such that the directed random walk X gets trapped in a leaf with probability 1, then

$$\liminf B_2(G_n) \;\geqslant\; B_2(G)\,.$$

Similarly, in Proposition A.18 we provides an easy-to-check sufficient condition ensuring that $\lim_n B_2(G_n) = B_2(G)$.

In the next section, we state our main continuity result (on which our study of the B_2 index of uniform leaf-labeled galled trees relies): that the B_2 index is essentially continuous for a generic class of models of random phylogenetic networks known as blowups of Galton–Watson trees.

3.3 Blowups of Galton–Watson trees

Informally, a blowup of a random rooted tree T is a random phylogenetic network that is obtained by: first, sampling T; then replacing each internal vertex v by an independent realization Γ_v of a random phylogenetic network (identifying v with the root of Γ_v, and each of the children of v with a leaf of Γ_v – see Figure 6). In this construction, we require that Γ_v depend on $d^+(v)$ only, where $d^+(v)$ denotes the number of children of v in T, and that the matching between the leaves of Γ_v and the children of v be chosen uniformly at random and independently of everything else.

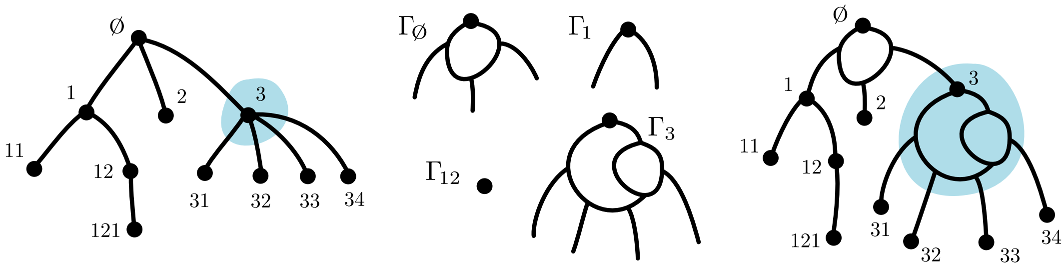

Figure 6: Illustration of the blowup: left, an ordered tree T with its Ulam–Harris labeling; center, the random networks associated to the internal vertices of T; right, the network G resulting from the blowup (with the Ulam–Harris labeling of the vertices of T indicated on the corresponding vertices of G, to make the link between T and G more apparent). Each vertex $v \in T$ has been "replaced" by the network Γ_v; for instance, vertex 3 corresponds to the subgraph of G highlighted in blue, which is isomorphic to Γ_3.

Formally, let T be a random ordered tree – that is, T is rooted and the children of each vertex $v \in T$ are ordered from 1 to $d^+(v)$ – which we view as a subset of the Ulam–Harris tree $\mathcal{U}$ (see e.g. [15, Section 6]). Let $\nu = (\nu_k)_{k \geqslant 1}$, where each ν_k is a leaf-exchangeable probability distribution on the set of finite phylogenetic networks with k leaves labeled $1, \ldots, k$ (by *leaf-exchangeable*, we mean invariant under permutation of the labels of the leaves). Finally, let $\Gamma = (\Gamma_v^k : v \in \mathcal{U}, k \geqslant 1)$ be a family of independent phylogenetic networks such that $\Gamma_v^k \sim \nu_k$ and Γ is independent of T.

The *blowup of T with respect to ν* is the random phylogenetic network obtained from (T, Γ) by gluing the networks $\Gamma_v := \Gamma_v^{d^+(v)}$, where v ranges over the internal vertices of T, as follows: for each internal vertex $v \in T$, let u_k denote the k-th child of v and, for each non-leaf u_k, identify the root of Γ_{u_k} with the leaf of Γ_v labeled k. Note that the leaves of the resulting network G are not properly labeled, but that they are in bijection with the leaves of T (which are canonically labeled by the Ulam–Harris labeling). To get a leaf-labeled network, one can label the leaves of T (e.g, uniformly at random) and carry over the labels to G; this is irrelevant when considering label-invariant functions of G, such at the B_2 index.

Blowups of Galton–Watson trees are a noteworthy class of random phylogenetic network. Indeed,

- Prominent "combinatorial" models of random phylogenetic networks – i.e. models that correspond to the uniform distribution on some relevant class of networks – can be obtained as blowups of (size-conditioned) Galton–Watson trees. For instance, in [26] Stufler proved that this is the case for uniform leaf-labeled level-k networks (and, therefore, for the galled trees considered here, as detailed in Section 3.4 below). Blowups of Galton–Watson trees can also result from biologically relevant evolutionary processes, see [2].
- The constrained structure of blowups of Galton–Watson trees makes them highly tractable. In [26], this was used to obtain asymptotic counting results for level-k networks, as as well as an explicit description of their large-scale geometry. Another example illustrating these constraints is given in [3], where it is proved that if a combinatorial model of phylogenetic networks can be obtained as a blowup up of a Galton–Watson tree, then for any fixed subgraph S the fraction of the networks that contain S in the corresponding combinatorial class is either 0 or 1.

The tractability of blowups of Galton–Watson trees comes, with varying degrees of directness, from that of the underlying trees. In particular, the remarkable asymptotic behavior of large size-conditioned critical Galton–Watson trees often plays a crucial role. Here, we only focus on their local limit – which, as we will see, has a universal structure known as *Kesten's tree.*

Definition 3.6. Let $\eta = (\eta_i)_{i \geqslant 0}$ be a probability distribution on the integers such that $\eta_0 > 0$ and $\sum_{i \geqslant 0} i\eta_i = 1$. The *Kesten tree associated to* η is the two-type (*spine*/*regular*) Galton–Watson tree T_* such that:

- *Regular* vertices have offspring distribution η, and all of their children are regular vertices.
- *Spine* vertices have offspring distribution $\hat{\eta}$ given by $\hat{\eta}_i = i\eta_i$, and exactly one of their children (whose order is chosen uniformly) is a spine vertex.
- The root of T_* is a spine vertex.

A Kesten tree T_* is always infinite, and has exactly one ray $v_0 \to v_1 \to \ldots$ starting from the root. This ray is known as the *spine.* ◇

Let T be a Galton–Watson tree with critical offspring distribution η such that $\eta_0 > 0$. For all n such that $\mathbb{P}(|T| = n) > 0$, where $|T|$ denotes the number of leaves of T, let T_n denote a random tree distributed as T conditioned to have n leaves. It is well-known that $T_n \to T_*$ in distribution for the local topology, where T_* is the Kesten tree associated to η; see e.g. [15, Theorem 7.1].

Let now G be a blowup of T with respect to some family ν of leaf-exchangeable finite random graphs, and let G_n be distributed as G conditioned to have n leaves, noting that G_n can equivalently be described as a blowup of T_n with respect to ν. Because the blowup procedure is a local operation, the convergence of T_n to T_* implies that $G_n \to G_*$ in distribution for the local topology, where G_* is a blowup of T_* with respect to ν.

The following theorem, which we prove in Appendix A.6, is a continuity criterion for the B_2 index of blowups of Galton–Watson trees.

Theorem 3.7. *With the notation above, assuming that the offspring distribution* η *is critical, satisfies* $\eta_0 > 0$ *and has a finite third moment, we have:*

(i) $B_2(G_n) \to B_2(G_*)$ *in distribution.*

(ii) *For all* $p \geqslant 1$, $\mathbb{E}[B_2(G_n)^p] \to \mathbb{E}[B_2(G_*)^p]$, *and all these moments are finite.*

Note the assumption in Theorem 3.7 that the offspring distribution η has a finite third moment. This is because, as already pointed out in Section 3.2, the mere convergence of G_n to G_* is not sufficient to get the convergence of $B_2(G_n)$: we also need to control the speed of convergence. Loosely speaking, we need an upper bound on the total variation distance between $[G_n]_{k_n}$ and $[G_*]_{k_n}$, for suitable sequences (k_n). This upper bound comes from a total variation bound for the convergence of Galton–Watson trees: see Proposition A.21 in Appendix A.5. This total variation bound relies on a Berry–Esseen type local central limit theorem, hence the third moment condition.

Combined with the independence between the base tree and the random networks used in the blowup procedure, the recursive structure of Galton–Watson trees makes it possible to get a simple expression for the expected value of the B_2 index.

Theorem 3.8. *Let* T *be a Galton–Watson whose offspring distribution* η *is such that* $\eta_0 > 0$, *and let* G *be a blowup of* T *with respect to* ν. *Then,*

$$\mathbb{E}[B_2(G)] \;=\; \frac{1}{\eta_0}\,\mathbb{E}[f(\xi)]\,,$$

where $\xi \sim \eta$, $f(0) = 0$ *and, for* $k \geqslant 1$, $f(k) = \mathbb{E}[B_2(\Gamma_k)]$, *where* $\Gamma_k \sim \nu_k$. *If in addition* η *has mean 1, then denoting by* T_* *the Kesten tree associated to* T *and by* G_* *the blowup of* T_* *with respect to* ν,

$$\mathbb{E}[B_2(G_*)] \;=\; \frac{1}{\eta_0}\Big(\mathbb{E}[f(\xi)] \;+\; \mathbb{E}[f(\hat{\xi}\,)]\Big)\,,$$

where $\hat{\xi} \sim \hat{\eta}$, *the size-biased distribution associated to* η *– that is,* $\hat{\eta}_k = k\,\eta_k$.

Proof. First, we use the recursive structure of G (see Figure 7) to characterize the distribution of $B_2(G)$. For this, note that:

- with probability η_0, G is reduced to its root, and $B_2(G) = 0$;
- with probability η_k, the root of T has k children and thus the network Γ associated with it is distributed as ν_k. In that case, by the grafting property,

$$B_2(G) \;\overset{d}{=}\; B_2(\Gamma) \;+\; \sum_{i=1}^{k} q_i\, B_2(G^{(i)})\,, \tag{11}$$

  where $G^{(1)}, \ldots, G^{(k)}$ are independent copies of G, and q_i is the probability that the directed random walk started from the root of Γ reaches its leaf labeled i.

Let ξ denote the number of children of the root of T, and write $\alpha := \mathbb{E}[B_2(G)]$ for conciseness. Conditional on $\{\xi = k\}$, we have that q_i is independent of $G^{(i)}$ and

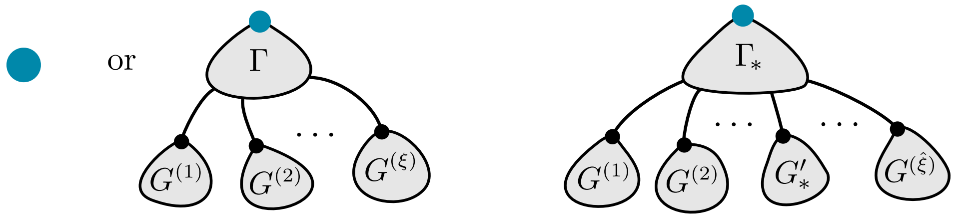

Figure 7: The recursive structure of G (on the left) and of G_* (on the right). The network G is either reduced to a single vertex or obtained by grafting ξ independent copies of itself on the leaves of a network $\Gamma \sim \nu_\xi$. By contrast, G_* is never reduced to a single vertex: it is obtained by grafting $\hat{\xi} - 1$ independent copies of G and an independent copy of itself on the leaves of a network $\Gamma_* \sim \nu_{\hat{\xi}}$.

that $\sum_{i=1}^{k} q_i = 1$. Moreover, each network $G^{(i)}$ is distributed as G. As a result, by taking expectations in Equation (11) we get

$$\begin{cases} \mathbb{E}(B_2(G) \,|\, \xi = 0) \;=\; 0 \\ \mathbb{E}(B_2(G) \,|\, \xi = k) \;=\; \alpha + f(k)\,, \end{cases}$$

where $f(k) := \mathbb{E}(B_2(\Gamma) \,|\, \xi = k)$. Plugging this in $\alpha = \sum_{k \geqslant 0} \eta_k \, \mathbb{E}(B_2(G) \,|\, \xi = k)$ and solving for α, we get $\mathbb{E}[B_2(G)] = \eta_0^{-1} \mathbb{E}[f(\xi)]$, proving the first part of the theorem.

Similarly, let $\hat{\xi}$ denote the number of leaves of the root of T_*, recalling that $\hat{\xi} \sim \hat{\eta}$, where $\hat{\eta}$ denotes the size-biased distribution associated to η. Let Γ_* denote the network associated to the root of T_*, and set $\alpha_* := \mathbb{E}[B_2(G_*)]$. As previously, write q_i for the probability that the directed random walk started from the root of Γ_* goes through its leaf labeled i. Finally, let S denote the leaf of Γ_* through which the spine of G_* goes, and write $I = \{1, \ldots, \hat{\xi}\} \setminus \{S\}$. Then,

$$B_2(G_*) \;\overset{d}{=}\; B_2(\Gamma_*) \;+\; \sum_{i \in I} q_i \, B_2(G^{(i)}) \;+\; q_S \, B_2(G'_*)$$

where $G^{(1)}, G^{(2)}, \ldots$ are independent copies of G, and G'_* is an independent copy of G_*. Taking expectations and recalling that $q_1 + \cdots + q_{\hat{\xi}} = 1$, we get

$$\alpha_* \;=\; \mathbb{E}[B_2(\Gamma_*)] \;+\; \alpha \;+\; (\alpha_* - \alpha)\, \mathbb{E}(q_S)\,, \tag{12}$$

In this expression, $\mathbb{E}[B_2(\Gamma_*)] = \mathbb{E}[f(\hat{\xi}\,)]$. Next, to compute $\mathbb{E}(q_S)$, note that

$$\mathbb{E}(q_S \,|\, \Gamma_*) \;=\; \sum_{i=1}^{\hat{\xi}} q_i \, \underbrace{\mathbb{P}(S = i \,|\, \Gamma_*)}_{1/\hat{\xi}} \;=\; \frac{1}{\hat{\xi}}\,,$$

from which it follows that $\mathbb{E}(q_S) = \sum_{k \geqslant 1} \frac{\hat{\eta}_k}{k} = \sum_{k \geqslant 1} \eta_k = 1 - \eta_0$. Plugging this in Equation (12) and solving for α_* then yields the desired expression. ☐

Note that since the function f in Theorem 3.8 satisfies $f(k) \leqslant \log_2 k$ for $k \geqslant 1$, this implies that if G is a blowup of T, then $\mathbb{E}[B_2(G)] \leqslant \mathbb{E}[B_2(T)]$. In fact, it is not too hard to see that we have the stronger statement

$$\mathbb{E}(B_2(G) \,|\, T) \;\leqslant\; B_2(T)\,,$$

as proved in Corollary A.27 from Appendix A.6. In particular, this immediately implies that if T is a random tree (not necessarily a Galton–Watson tree) and G is a blowup of T, then $B_2(T)$ is second-order stochastically dominant over $B_2(G)$; see Corollary A.28.

Having introduced all the required tools, we now turn to the study of the B_2 index of uniform leaf-labeled galled trees.

3.4 Application to leaf-labeled galled trees

Let us start by briefly recalling the blowup construction of uniform leaf-labeled galled trees introduced by Stufler in [26]; we refer the reader to this article for a rigorous justification of the sampling procedure outlined below.

Following [26], let a *head of size* k be:

- for $k = 2$, a galled tree with 2 labeled leaves;
- for $k \geqslant 3$, a galled tree with a single gall of length $k+1$ and k labeled leaves.

Let then a_k denote the number of heads of size k – so that $a_1 = 0$, $a_2 = 3$, $a_3 = 9$; see Figure 8. To count the number of heads of size $k \geqslant 3$: first, orient the gall and choose the position of the reticulation in the gall (k possibilities); then, label the k leaves ($k!$ possibilities); finally, forget the orientation of the cycle, and observe that each head is obtained exactly twice. This reasoning also works for $k = 2$, but in that case we have to add the network consisting of a single cherry. Putting the pieces together, we get $a_1 = 0$ and, for $k \geqslant 2$,

$$a_k \;=\; \frac{k \cdot k!}{2} \;+\; \mathbb{1}_{\{k=2\}}\,.$$

Next, let $\nu = (\nu_k)_{k\geqslant 2}$, where ν_k denotes the uniform distribution on the heads of size k. It is not too hard to see that if T_n is a simply generated tree with n leaves (see e.g. [15]) whose weight sequence is given by $w_k = a_k/k!$, then the blowup of T_n with respect to ν is a uniform leaf-labeled galled tree.

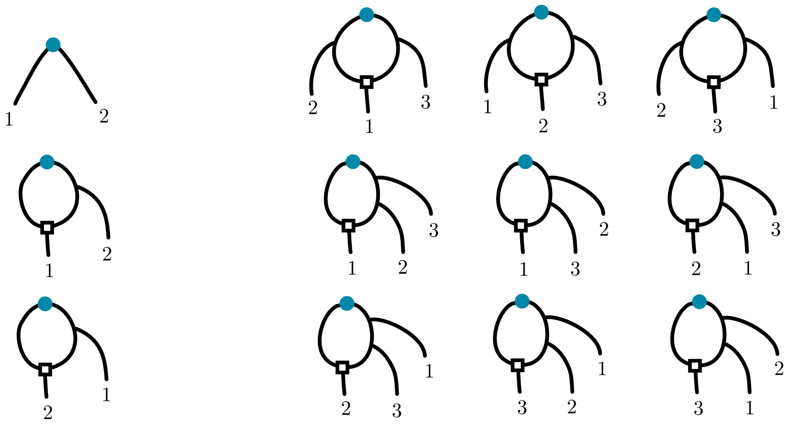

Figure 8: The heads of size k used in the blowup construction of galled trees, for $k = 2$ (left) and $k = 3$ (right). The blue dots are the roots, and the white squares are reticulations. Note that the heads are leaf-labeled.

To show that the simply generated tree T_n can also be obtained as a conditioned critical Galton–Watson tree, we must find an offspring distribution $(\eta_k)_{k\geqslant 0}$ with mean 1 that is an exponential tilt of the weight sequence $(w_k)_{k\geqslant 1}$ – that is, we must find $\theta > 0$ such that

$$\begin{cases} \eta_k = \theta^{k-1}\, w_k \ \text{ for } k \geqslant 1 \text{ and } \eta_0 = 1 - \sum_{k\geqslant 1} \eta_k & \text{(exponential tilt)} \\ \sum_{k\geqslant 0} k\, \eta_k = 1 & \text{(criticality).} \end{cases}$$

Note that the specific form of this exponential tilt (where, to get a probability distribution, we change η_0 – instead of by multiplying all η_k's by a constant) is due to the fact that we are conditioning on the number of leaves (as opposed to the number of vertices).

To find θ, set $g(z) = \sum_{k\geqslant 1} w_k\, z^k$, so that $\sum_{k\geqslant 1} k\, \eta_k = 1 \iff g'(\theta) = 1$. Elementary calculations then show that

$$g'(z) \;=\; \frac{(2z^2 - 3z + 3)(z-2)z}{2(z-1)^3}\,,$$

from which we get

$$\theta \;=\; \frac{5 - \sqrt{17}}{4}\,.$$

Note in passing that θ is also the first-order term in the expansion of the generating function of galled trees at $z = 1/8$ – see Eq. (2).

Let us recap the construction.

Proposition 3.9. *Let T_n be a critical Galton–Watson tree conditioned to have n leaves whose offspring distribution $(\eta_k)_{k\geqslant 0}$ is defined by*

$$\eta_0 = \frac{(2-\theta)(\theta^2 - 3\theta + 1)}{2(\theta-1)^2}, \qquad \eta_1 = 0, \qquad \eta_2 = \frac{3}{2}\,\theta, \qquad \eta_k = \frac{k}{2}\,\theta^{k-1}\ (k \geqslant 3),$$

with $\theta = (5-\sqrt{17})/4$. Let then G_n be the blowup of T_n with respect to $(\nu_k)_{k\geqslant 2}$, where ν_k denotes the uniform distribution on the set of heads of size k. Then, G_n is uniformly distributed on the set of galled trees with n labeled leaves.

Since the offspring distribution η in Proposition 3.9 has a finite third moment, it follows from Theorem 3.7 that, as $n \to \infty$, $B_2(G_n)$ converges in distribution and in all moments to $B_2(G_*)$, where G_* is the local limit of G_n – which, as we have already seen, is the blowup with respect to ν of the Kesten tree T_* associated to η. Therefore, $\lim_n \mathbb{E}[B_2(G_n)] = \mathbb{E}[B_2(G_*)]$ can be computed using the expression given in Theorem 3.8.

Because when $\xi = 2$ the head can be a cherry (whose structure differs from other heads, see Figure 8), it will be convenient to let β_k denote the expected B_2 index of a network Γ_k sampled uniformly at random among the galled trees with a single gall of length $k+1$ and k labeled leaves; and to write the function f from Theorem 3.8 as $f(2) = \frac{1}{3} + \frac{2}{3}\beta_2$ for $k = 2$ and $f(k) = \beta_k$ for $k \geqslant 3$.

To compute β_k, orient the gall of Γ_k uniformly at random, then number its vertices from 1 to k. For $r = 0, \ldots, k-1$, let $\Gamma_{k,r}$ denote the realization of Γ_k whose reticulation is on the $(r+1)$-th vertex of the gall – that is, there are r vertices on one side of the gall and $k-r-1$ on the other. Letting $\beta_{k,r} = B_2(\Gamma_{k,r})$, we have

$$\beta_{k,r} \;=\; \sum_{i=1}^{r} \frac{i+1}{2^{i+1}} \;+\; \sum_{j=1}^{k-r-1} \frac{j+1}{2^{j+1}} \;-\; \left(\frac{1}{2^{r+1}} + \frac{1}{2^{k-r}}\right) \log_2\left(\frac{1}{2^{r+1}} + \frac{1}{2^{k-r}}\right)\,.$$

Next, we need to integrate over the position of the reticulation, as described by r in the parametrization above. Using the procedure described at the beginning of the section to enumerate the heads, we see that r is uniform on $\{0, \ldots, k-1\}$. Therefore,

$$\beta_k \;=\; \frac{1}{k}\sum_{r=0}^{k-1} \beta_{k,r}\,. \tag{13}$$

Finally, by substituting the expression for η_k from Proposition 3.9 in the expression of Theorem 3.8, we get – in the notation of the theorem,

$$\mathbb{E}[f(\xi)] \;=\; \frac{\theta}{2} \;+\; \frac{1}{2}\sum_{k\geqslant 2} k\,\theta^{k-1}\,\beta_k \quad \text{and} \quad \mathbb{E}[f(\hat{\xi}\,)] \;=\; \theta \;+\; \frac{1}{2}\sum_{k\geqslant 2} k^2\,\theta^{k-1}\beta_k\,.$$

The following theorem summarizes the calculations of this section; compare with the results from Section 2.

Theorem 3.10. *Let G_n be sampled uniformly at random among the set of galled trees with n labeled leaves. Then, as $n \to \infty$, $B_2(G_n)$ converges in distribution and in all moments to a random variable $B_2(G_*)$ with finite moments of all orders. In particular,*

$$\mathbb{E}(B_2(G_*)) \;=\; \frac{1}{2\,\eta_0}\left(3\,\theta \;+\; \sum_{k\geqslant 2}(k+1)k\,\theta^{k-1}\beta_k\right),$$

where θ and η_0 are explicit algebraic constants given in Proposition 3.9, and β_k is given in Equation (13). Numerically, $\alpha_* = 2.707911858984\ldots$.

Acknowledgments

Part of this work was done while FB was at the Institute for Theoretical Studies of ETH Zürich and was supported by Dr. Max Rössler, the Walter Haefner Foundation and the ETH Zürich Foundation. MF and TCY acknowledge partial support by National Science and Technology Council (NSTC), Taiwan under grant NSTC-111-2115-M-004-002-MY2.

References

[1] The On-Line Encyclopedia of Integer Sequences, published electronically at `http://oeis.org`, 2024.

[2] F. Bienvenu and J.-J. Duchamps. A branching process with coalescence to model random phylogenetic networks. *Electronic Journal of Probability*, 29: 1–48, 2024. `DOI:10.1214/24-EJP1088`.

[3] F. Bienvenu and M. Steel. 0–1 laws for pattern occurrences in phylogenetic trees and networks. *Bulletin of Mathematical Biology*, 86(94), 2024. `DOI: 10.1007/s11538-024-01316-x`.

[4] F. Bienvenu, G. Cardona, and C. Scornavacca. Revisiting Shao and Sokal's B_2 index of phylogenetic balance. *Journal of Mathematical Biology*, 83(52): 1–43, 2021. `DOI:10.1007/s00285-021-01662-7`.

[5] P. Billingsley. *Probability and Measure.* John Wiley & Sons, 3rd edition, 1995. ISBN 978-0-471-00710-4.

[6] P. Billingsley. *Convergence of Probability Measures.* John Wiley & Sons, 2nd edition, 1999. `DOI:10.1002/9780470316962`.

[7] M. Bouvel, P. Gambette, and M. Mansouri. Counting phylogenetic networks of level 1 and 2. *Journal of Mathematical Biology*, 81(6):1357–1395, 2020. `DOI:10.1007/s00285-020-01543-5`.

[8] G. Cardona, M. Llabrés, F. Rosselló, and G. Valiente. Comparison of galled trees. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(2):410–427, 2010. `DOI:10.1109/TCBB.2010.60`.

[9] M. Dwass. The total progeny in a branching process and a related random walk. *Journal of Applied Probability*, 6(3):682–686, 1969. `DOI:10.2307/3212112`.

[10] M. Fischer, L. Herbst, S. Kersting, A. L. Kühn, and K. Wicke. *Tree Balance Indices: A Comprehensive Survey.* Springer, Cham, 2023. `DOI:10.1007/978-3-031-39800-1`.

[11] P. Flajolet and R. Sedgewick. *Analytic Combinatorics.* Cambridge University Press, 2009. `DOI:10.1017/CBO9780511801655`.

[12] M. Fuchs and B. Gittenberger. Sackin indices for labeled and unlabeled classes of galled trees. *arXiv preprint*, 2024. `arXiv:2407.13892`.

[13] C. C. Heyde. On large deviation problems for sums of random variables which are not attracted to the normal law. *The Annals of Mathematical Statistics*, 38(5):1575–1578, 1967. `DOI:10.1214/aoms/1177698712`.

[14] D. H. Huson, R. Rupp, and C. Scornavacca. *Phylogenetic Networks: Concepts, Algorithms and Applications.* Cambridge University Press, 2010. `DOI:10.1017/CBO9780511974076`.

[15] S. Janson. Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation. *Probability Surveys*, 9:103–252, 2012. `DOI:10.1214/11-PS188`.

[16] O. Kallenberg. *Foundations of Modern Probability.* Springer Cham, 3 edition, 2021. `DOI:10.1007/978-3-030-61871-1`.

[17] J. G. Kemeny, J. L. Snell, and A. W. Knapp. *Denumerable Markov Chains.* Springer New York, NY, 2nd edition, 1976. `DOI:10.1007/978-1-4684-9455-6`.

[18] G. Kersting. On the height profile of a conditioned Galton–Watson tree. *arXiv*, 2011 (first version dates from 1998). `arXiv:1101.3656`.

[19] L. Knüver, M. Fischer, M. Hellmuth, and K. Wicke. The weighted total cophenetic index: A novel balance index for phylogenetic networks. *arXiv preprint*, 2023. `arXiv:2307.08654`.

[20] I. Kortchemski. Invariance principles for Galton–Watson trees conditioned on the number of leaves. *Stochastic Processes and their Applications*, 122(9): 3126–3172, 2012. `DOI:10.1016/j.spa.2012.05.013`.

[21] N. F. Martin and J. W. England. *Mathematical Theory of Entropy*. Cambridge University Press, 1984. `DOI:10.1017/CBO9781107340718`.

[22] N. Minami. On the number of vertices with a given degree in a Galton-Watson tree. *Advances in Applied Probability*, 37(1):229–264, 2005. `DOI:10.1239/aap/1113402407`.

[23] V. V. Petrov. *Sums of Independent Random Variables.* Springer, Berlin, Heidelberg, 1975. `DOI:10.1007/978-3-642-65809-9`.

[24] K.-T. Shao and R. R. Sokal. Tree balance. *Systematic Zoology*, 39(3):266–276, 1990. `DOI:10.2307/2992186`.

[25] B. Stufler. Local limits of large Galton–Watson trees rerooted at a random vertex. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 55 (1):155–183, 2019. `DOI:10.1214/17-AIHP879`.

[26] B. Stufler. A branching process approach to level-k phylogenetic networks. *Random Structures & Algorithms*, 61(2):397–421, 2022. `DOI:10.1002/rsa.21065`.

[27] R. van der Hofstad. *Random Graphs and Complex Networks*, volume 2. Cambridge University Press, 2024. `DOI:10.1017/9781316795552`.

[28] W. Woess. *Denumerable Markov Chains.* European Mathematical Society, 2009. `DOI:10.4171/071`.

[29] L. Zhang. The Sackin index of simplex networks. In *Comparative Genomics: 19th International Conference, RECOMB-CG 2022*, pages 52–67. Springer, Cham, 2022. `DOI:10.1007/978-3-031-06220-9_4`.

Appendices

A.1 The limit law via the method of moments

The main purpose of this appendix is to prove the technical estimate (10) (see Lemma A.2 below) in order to complete the proof of Theorem 2.3. Set

$$f_{\ell,1}(z) := \frac{1}{2}\sum_{\substack{\ell_1+\ell_2+\ell_3=\ell\\ \ell_1,\ell_2<\ell}} \binom{\ell}{\ell_1,\ell_2,\ell_3}\frac{1}{2^{\ell_1+\ell_2}}A^{[\ell_1]}(z)A^{[\ell_2]}(z),$$

$$f_{\ell,2}(z) := \sum_{k\geqslant 1}\sum_{\substack{\ell_1+\cdots+\ell_{k+2}=\ell\\ \ell_j<\ell,1\leqslant j\leqslant k+1}} \binom{\ell}{\ell_1,\ldots,\ell_{k+2}}\prod_{s=1}^{k}\frac{1}{2^{(s+1)\ell_s}}\left(\frac{1}{2}+\frac{1}{2^{k+1}}\right)^{\ell_{k+1}}\mu_{k,0}^{\ell_{k+2}} \times\prod_{s=1}^{k+1}A^{[\ell_s]}(z),$$

$$f_{\ell,3}(z) := \frac{1}{2}\sum_{k,m\geqslant 1}\sum_{\substack{\ell_1+\cdots+\ell_{k+m+2}=\ell\\ \ell_j<\ell,1\leqslant j\leqslant k+m+1}} \binom{\ell}{\ell_1,\cdots,\ell_{k+m+2}}\prod_{s=1}^{k}\frac{1}{2^{(s+1)\ell_s}}\prod_{t=1}^{m}\frac{1}{2^{(t+1)\ell_{k+t}}} \times\left(\frac{1}{2^{k+1}}+\frac{1}{2^{m+1}}\right)^{\ell_{k+m+1}}\mu_{k,m}^{\ell_{k+m+2}}\prod_{s=1}^{k}A^{[\ell_s]}(z)\prod_{t=1}^{m+1}A^{[\ell_{k+t}]}(z)$$

so that $f_\ell(z) = f_{\ell,1}(z) + f_{\ell,2}(z) + f_{\ell,3}(z)$; see (6), where we have split the last sum of (6) into two parts since below, we use the first (simpler) part to explain our ideas and then treat the second (more complicated) part without repeating the similar details. Moreover, set

$$c_{\ell,i} := f_{\ell,i}(A^{[j]}(z) \leftrightarrow c_j), \qquad d_{\ell,i} := f'_{\ell,i}((A^{[j]})'(z) \leftrightarrow d_j, A^{[j]}(z) \leftrightarrow c_j),$$

and thus

$$c_\ell = \frac{c_{\ell,1}+c_{\ell,2}+c_{\ell,3}}{g_\ell(\rho)}, \qquad d_\ell = \frac{d_{\ell,1}+d_{\ell,2}+d_{\ell,3}}{g_\ell(\rho)} - c_\ell\frac{g'_\ell(\rho)}{g_\ell(\rho)}, \tag{14}$$

where notation is as in Section 2.

We first consider c_ℓ for which we have the following result.

Lemma A.1. *There exists a constant $K > 0$ such that $c_\ell \leqslant K^\ell \ell!$ for all $\ell \geqslant 0$.*

Proof. We prove the claim by induction on ℓ, where we take

$$K := \max\left\{c_1, \sqrt{\frac{c_2}{2}}, \ldots, \sqrt[9]{\frac{c_9}{9!}}, 512\right\}.$$

Note that by this choice, the claim holds for $1 \leqslant \ell \leqslant 9$. In addition, it trivially holds for $\ell = 0$. Also, note that $g_\ell(\rho)$ is increasing in ℓ and $g_1(\rho) \equiv 0.570194101601\cdots$. Thus, by (14), we have to show that

$$c_{\ell,1} + c_{\ell,2} + c_{\ell,3} \leqslant g_1(\rho)K^\ell \ell!$$

for $\ell \geqslant 10$. We assume that this claim holds for $\ell' < \ell$ and prove it for ℓ. For this, we estimate $c_{\ell,i}$ with $i = 1, 2, 3$ separately.

First, for $c_{\ell,1}$, we have:

$$\begin{aligned}
c_{\ell,1} &\leqslant \frac{1}{2}\sum_{\substack{\ell_1+\ell_2+\ell_3=\ell\\ \ell_1,\ell_2<\ell}} \frac{\ell!}{\ell_3!}\frac{1}{2^{\ell_1+\ell_2}}K^{\ell_1+\ell_2} \leqslant \frac{\ell!}{2}\sum_{\substack{\ell_1+\ell_2+\ell_3=\ell\\ \ell_1,\ell_2<\ell}} \frac{K^{\ell-\ell_3}}{2^{\ell-\ell_3}}\\
&= \frac{\ell!}{2}\left(\frac{K}{2}\right)^{\ell}\sum_{\substack{\ell_1+\ell_2+\ell_3=\ell\\ \ell_1,\ell_2<\ell}}\left(\frac{2}{K}\right)^{\ell_3} \leqslant \frac{\ell!}{2}\left(\frac{K}{2}\right)^{\ell}\sum_{\ell_3=0}^{\ell}\left(\frac{1}{256}\right)^{\ell_3}\sum_{\substack{\ell_1+\ell_2=\ell-\ell_3\\ \ell_1,\ell_2<\ell}} 1\\
&\leqslant K^{\ell}\ell!\left(\frac{1}{2}\right)^{\ell+1}\sum_{\ell_3=0}^{\ell}\left(\frac{1}{256}\right)^{\ell_3}(\ell-\ell_3+1)\\
&= K^{\ell}\ell!\left(\frac{1}{2}\right)^{\ell+1}\left(\frac{256}{255}\ell+\frac{256^{-\ell}}{65025}+\frac{65024}{65025}\right)\\
&\leqslant 0.005390234525\cdots K^{\ell}\ell!, \qquad (15)
\end{aligned}$$

where in the estimate of the second line, we used that $K \geqslant 512$, and the last inequality holds for $\ell \geqslant 10$.

Next, we consider $c_{\ell,2}$. Here, we first note that since $\ell_1+\cdots+\ell_{k+2}=\ell$, we have $k+1$ degrees of freedom, i.e., ℓ_{k+1} is fixed if $\ell_1,\ldots,\ell_k$ and ℓ_{k+2} are decided. We replace $c_{\ell_{k+1}}$ on the right-hand side of $c_{\ell,2}$ by $K^{\ell_{k+1}}\ell_{k+1}!$ except when $\ell_{k+1}=0$ and $\ell_{k+1}=1$ where we replace it by ρ and c_1, respectively. This gives

$$c_{\ell,2} \leqslant \frac{9}{16}\sum_{k\geqslant 1}{\sum}'\binom{\ell}{\ell_1,\ldots,\ell_{k+2}}\prod_{s=1}^{k}\frac{1}{2^{(s+1)\ell_s}}\mu_{k,0}^{\ell_{k+2}}\left(\prod_{s=1}^{k}c_{\ell_s}\right)K^{\ell_{k+1}}\ell_{k+1}!, \qquad (16)$$

where $\ell_{k+1}=\ell-\ell_1-\cdots-\ell_k-\ell_{k+2}$, the second sum is over all $0<\ell_1+\cdots+\ell_k+\ell_{k+2}\leqslant \ell$ with $\ell_j<\ell, 1\leqslant j\leqslant k$, and we have used that $\xi_{k,\ell_{k+1}}\leqslant 9/16$ with

$$\xi_{k,\ell_{k+1}} = \begin{cases} \rho, & \text{if } \ell_{k+1}=0,\\ \left(\frac{1}{2}+\frac{1}{2^{k+1}}\right)\frac{c_1}{K}, & \text{if } \ell_{k+1}=1,\\ \left(\frac{1}{2}+\frac{1}{2^{k+1}}\right)^{\ell_{k+1}}, & \text{if } \ell_{k+1}\geqslant 2.\end{cases}$$

Next, by using $\mu_{k,0}\leqslant 2$ and plugging the induction hypothesis for c_{ℓ_s} into (16) except when $s\neq k+1$ where we again use ρ and c_1 if $\ell_s=0$ and $\ell_s=1$, we have

$$\begin{aligned}
c_{\ell,2} &\leqslant \frac{9}{16}K^{\ell}\ell!\sum_{k\geqslant 1}\prod_{s=1}^{k}\left(\rho+\frac{c_1}{K2^{s+1}}+\sum_{\ell_s\geqslant 2}\frac{1}{2^{(s+1)\ell_s}}\right)\sum_{\ell_{k+2}=0}^{\infty}\frac{2^{\ell_{k+2}}}{K^{\ell_{k+2}}\ell_{k+2}!}\\
&\leqslant \frac{9e^{1/256}}{16}K^{\ell}\ell!\sum_{k\geqslant 1}\prod_{s=1}^{k}\left(\rho+\frac{c_1}{K2^{s+1}}+\frac{1}{2^{s+1}(2^{s+1}-1)}\right)\\
&= 0.223033369804\cdots K^{\ell}\ell!, \qquad (17)
\end{aligned}$$

where we used that $K\geqslant 512$ in the second last step and Maple to evaluate the numerical constant.

Similar, we obtain for $c_{\ell,3}$ by using $\mu_{k,m}\leqslant 3$ that

$$c_{\ell,3}\leqslant \frac{\rho e^{3/512}}{2}K^{\ell}\ell!\sum_{k\geqslant 1}\prod_{s=1}^{k}\left(\rho+\frac{c_1}{K2^{s+1}}+\frac{1}{2^{s+1}(2^{s+1}-1)}\right)$$

$$\times \sum_{m\geqslant 1}\prod_{t=1}^{m}\left(\rho+\frac{c_1}{K2^{t+1}}+\frac{1}{2^{t+1}(2^{t+1}-1)}\right)$$
$$= 0.017199020110\cdots K^{\ell}\ell!. \tag{18}$$

Combining (15), (17), and (18) gives

$$c_{\ell,1}+c_{\ell,2}+c_{\ell,3}\leqslant 0.245622624440\cdots K^{\ell}\ell!\leqslant g_1(\rho)K^{\ell}\ell!$$

which proves the desired result. □

We use this now to prove a similar result for d_ℓ.

Lemma A.2. *There exists a constant $K>0$ such that $d_\ell\leqslant K^{\ell}\ell!$ for all $\ell\geqslant 0$.*

Proof. We again use induction on ℓ where the claim holds for $\ell=0$. Moreover, we choose K large enough such that the conclusion of Lemma A.1 holds and in addition that

(i) $K\geqslant \max\left\{d_1,\sqrt{d_2/2},\dots,\sqrt{d_9/9!},4608c'/c_1,4608d'/d_1\right\}$;

(ii) $\rho\geqslant \max\left\{3c_1/(4K),3^2c_2/(2!4^2K^2),\dots,3^9c_9/(9!4^9K^9)\right\}$,

where $c':=\max\{c_1,\dots,c_9\}$ and $d':=\max\{d_1,\dots,d_9\}$.

By the item (i), the claim holds for all $1\leqslant\ell\leqslant 9$. Thus, we can assume that $\ell\geqslant 10$ and that the claim holds for all $\ell'<\ell$. We are going to bound $d_{\ell,i}$ for $i=1,2,3$.

First for $d_{\ell,1}$, we have

$$f'_{\ell,1}(z)=\frac{1}{2}\sum_{\substack{\ell_1+\ell_2+\ell_3=\ell\\ \ell_1,\ell_2<\ell}}\binom{\ell}{\ell_1,\ell_2,\ell_3}\frac{1}{2^{\ell_1+\ell_2}}\left((A^{[\ell_1]})'(z)A^{[\ell_2]}(z)+A^{[\ell_1]}(z)(A^{[\ell_2]})'(z)\right)$$

and thus

$$\begin{aligned}d_{\ell,1}&=\frac{1}{2}\sum_{\substack{\ell_1+\ell_2+\ell_3=\ell\\ \ell_1,\ell_2<\ell}}\binom{\ell}{\ell_1,\ell_2,\ell_3}\frac{1}{2^{\ell_1+\ell_2}}\left(d_{\ell_1}c_{\ell_2}+c_{\ell_1}d_{\ell_2}\right)\\ &=\sum_{\substack{\ell_1+\ell_2+\ell_3=\ell\\ \ell_1,\ell_2<\ell}}\binom{\ell}{\ell_1,\ell_2,\ell_3}\frac{1}{2^{\ell_1+\ell_2}}c_{\ell_1}d_{\ell_2}.\end{aligned}$$

Using the induction hypothesis and the estimate from Lemma A.1, we obtain

$$d_{\ell,1}\leqslant\sum_{\substack{\ell_1+\ell_2+\ell_3=\ell\\ \ell_1,\ell_2<\ell}}\frac{\ell!}{\ell_3!}\frac{1}{2^{\ell_1+\ell_2}}K^{\ell_1+\ell_2}\leqslant 0.010780469050\cdots K^{\ell}\ell! \tag{19}$$

which is twice the bound from (15).

Next, for $d_{\ell,2}$, we have

$$f'_{\ell,2}(z)=\sum_{k\geqslant 1}\sum_{\substack{\ell_1+\cdots+\ell_{k+2}=\ell\\ \ell_j<\ell,1\leqslant j\leqslant k+1}}\binom{\ell}{\ell_1,\dots,\ell_{k+2}}\prod_{s=1}^{k}\frac{1}{2^{(s+1)\ell_s}}\left(\frac{1}{2}+\frac{1}{2^{k+1}}\right)^{\ell_{k+1}}\mu_{k,0}^{\ell_{k+2}}$$

$$\times\left(\sum_{i=1}^{k+1}(A^{[\ell_i]})'(z)\prod_{\substack{1\leqslant j\leqslant k+1\\ j\neq i}}A^{[\ell_j]}(z)\right)$$

and thus

$$\begin{aligned}
d_{\ell,2} &= \sum_{k\geqslant 1}\sum_{\substack{\ell_1+\cdots+\ell_{k+2}=\ell\\ \ell_j<\ell,1\leqslant j\leqslant k+1}}\binom{\ell}{\ell_1,\cdots,\ell_{k+2}}\prod_{s=1}^{k}\frac{1}{2^{(s+1)\ell_s}}\left(\frac{1}{2}+\frac{1}{2^{k+1}}\right)^{\ell_{k+1}}\mu_{k,0}^{\ell_{k+2}}\\
&\qquad\times\left(\sum_{i=1}^{k+1}d_{\ell_i}\prod_{\substack{j=1\\ j\neq i}}^{k+1}c_{\ell_j}\right)\\
&= \sum_{k\geqslant 1}\sum_{\substack{\ell_1+\cdots+\ell_{k+2}=\ell\\ \ell_j<\ell,1\leqslant j\leqslant k+1}}\binom{\ell}{\ell_1,\cdots,\ell_{k+2}}\prod_{s=1}^{k}\frac{1}{2^{(s+1)\ell_s}}\left(\frac{1}{2}+\frac{1}{2^{k+1}}\right)^{\ell_{k+1}}\mu_{k,0}^{\ell_{k+2}}\\
&\qquad\times\left(\sum_{i=1}^{k}d_{\ell_i}\prod_{\substack{j=1\\ j\neq i}}^{k+1}c_{\ell_j}\right)\\
&\quad+\sum_{k\geqslant 1}\sum_{\substack{\ell_1+\cdots+\ell_{k+2}=\ell\\ \ell_j<\ell,1\leqslant j\leqslant k+1}}\binom{\ell}{\ell_1,\cdots,\ell_{k+2}}\prod_{s=1}^{k}\frac{1}{2^{(s+1)\ell_s}}\left(\frac{1}{2}+\frac{1}{2^{k+1}}\right)^{\ell_{k+1}}\mu_{k,0}^{\ell_{k+2}}\\
&\qquad\times d_{\ell_{k+1}}\prod_{j=1}^{k}c_{\ell_j}.
\end{aligned}$$

We use now both the estimate for c_ℓ from Lemma A.1 and the induction hypothesis for d_ℓ except for $0\leqslant\ell\leqslant 9$. Moreover, we use $\mu_{k,0}\leqslant 2$ (as in the proof of Lemma A.1) and

$$\left(\frac{1}{2}+\frac{1}{2^{k+1}}\right)^{\ell_{k+1}}c_{\ell_{k+1}}\leqslant\left(\frac{3}{4}\right)^{\ell_{k+1}}c_{\ell_{k+1}}\leqslant\rho K^{\ell_{k+1}}\ell_{k+1}!$$

which follows by item (ii) above and

$$\left(\frac{1}{2}+\frac{1}{2^{k+1}}\right)^{\ell_{k+1}}d_{\ell_{k+1}}\leqslant K^{\ell_{k+1}}\ell_{k+1}!$$

which holds trivially. The rest is handled by similar ideas as in the proof of Lemma A.1. Thus, we obtain

$$\begin{aligned}
d_{\ell,2} &\leqslant \rho e^{2/K}K^\ell\ell!\sum_{k\geqslant 1}\prod_{s=1}^{k}\left(\rho+\left(\sum_{i=1}^{9}\frac{c_i}{2^{i(s+1)}i!K^i}\right)+\frac{1}{2^{9(s+1)}(2^{s+1}-1)}\right)\\
&\qquad\times\left(\sum_{s=1}^{k}\frac{1+\left(\sum_{i=1}^{9}\frac{d_i}{2^{i(s+1)}i!K^i}\right)+\frac{1}{2^{9(s+1)}(2^{s+1}-1)}}{\rho+\left(\sum_{i=1}^{9}\frac{c_i}{2^{i(s+1)}i!K^i}\right)+\frac{1}{2^{9(s+1)}(2^{s+1}-1)}}\right)\\
&\quad+e^{2/K}K^\ell\ell!\sum_{k\geqslant 1}\prod_{s=1}^{k}\left(\rho+\left(\sum_{i=1}^{9}\frac{c_i}{2^{i(s+1)}i!K^i}\right)+\frac{1}{2^{9(s+1)}(2^{s+1}-1)}\right).
\end{aligned}$$

Note that, by our choice of K,

$$\sum_{i=1}^{9}\frac{c_i}{2^{i(s+1)}i!K^i}\leqslant\frac{9c'}{2^{s+1}K}\leqslant\frac{1}{512\cdot 2^{s+1}}$$

and likewise with the c_i's replaced by d_i's. Plugging this into the above expression and numerical evaluating it (again with Maple), we obtain

$$d_{\ell,2} \leqslant 0.643482769458 \cdots K^{\ell} \ell!. \tag{20}$$

We finally consider $d_{\ell,3}$ which is treated in a similar fashion. First,

$$\begin{aligned} f'_{\ell,3}(z) = \frac{1}{2} \sum_{k\geqslant 1} \sum_{m\geqslant 1} \sum_{\substack{\ell_1+\cdots+\ell_{k+m+2}=\ell \\ \ell_j<\ell, 1\leqslant j\leqslant k+m+1}} &\binom{\ell}{\ell_1, \cdots, \ell_{k+m+2}} \prod_{s=1}^{k} \frac{1}{2^{(s+1)\ell_s}} \prod_{t=1}^{m} \frac{1}{2^{(t+1)\ell_{k+t}}} \\ &\times \left(\frac{1}{2^{k+1}} + \frac{1}{2^{m+1}}\right)^{\ell_{k+m+1}} \mu_{k,m}^{\ell_{k+m+2}} \left(\sum_{i=1}^{k+m+1} (A^{[\ell_i]})'(z) \prod_{\substack{j=1 \\ j\neq i}}^{k+m+1} A^{[\ell_j]}(z) \right) \end{aligned}$$

and thus

$$\begin{aligned} d_{\ell,3} = \frac{1}{2} \sum_{k\geqslant 1} \sum_{m\geqslant 1} \sum_{\substack{\ell_1+\cdots+\ell_{k+m+2}=\ell \\ \ell_j<\ell, 1\leqslant j\leqslant k+m+1}} &\binom{\ell}{\ell_1, \cdots, \ell_{k+m+2}} \prod_{s=1}^{k} \frac{1}{2^{(s+1)\ell_s}} \prod_{t=1}^{m} \frac{1}{2^{(t+1)\ell_{k+t}}} \\ &\times \left(\frac{1}{2^{k+1}} + \frac{1}{2^{m+1}}\right)^{\ell_{k+m+1}} \mu_{k,m}^{\ell_{k+m+2}} \left(\sum_{i=1}^{k+m} d_{\ell_i} \prod_{\substack{j=1 \\ j\neq i}}^{k+m+1} c_{\ell_j} \right) \\ + \frac{1}{2} \sum_{k\geqslant 1} \sum_{m\geqslant 1} \sum_{\substack{\ell_1+\cdots+\ell_{k+m+2}=\ell \\ \ell_j<\ell, 1\leqslant j\leqslant k+m+1}} &\binom{\ell}{\ell_1, \cdots, \ell_{k+m+2}} \prod_{s=1}^{k} \frac{1}{2^{(s+1)\ell_s}} \prod_{t=1}^{m} \frac{1}{2^{(t+1)\ell_{k+t}}} \\ &\times \left(\frac{1}{2^{k+1}} + \frac{1}{2^{m+1}}\right)^{\ell_{k+m+1}} \mu_{k,m}^{\ell_{k+m+2}} d_{\ell_{k+m+1}} \prod_{i=1}^{k+m} c_{\ell_i}(z). \end{aligned}$$

Now, using $\mu_{k,m} \leqslant 3$ (see the proof of Lemma A.1) and arguments as above, we obtain

$$\begin{aligned} d_{\ell,3} &\leqslant \rho e^{3/K} K^{\ell} \ell! \sum_{k\geqslant 1} \prod_{s=1}^{k} \left(\rho + \frac{c_1}{512 \cdot 2^{(s+1)} K} + \frac{1}{2^{9(s+1)}(2^{s+1}-1)}\right) \\ &\times \left(\sum_{s=1}^{k} \frac{1 + \frac{d_1}{512\cdot 2^{(s+1)}} + \frac{1}{2^{9(s+1)}(2^{s+1}-1)}}{\rho + \frac{c_1}{512\cdot 2^{(s+1)}} + \frac{1}{2^{9(s+1)}(2^{s+1}-1)}} \right) \sum_{k\geqslant 1} \prod_{s=1}^{k} \left(\rho + \frac{c_1}{512 \cdot 2^{(s+1)}} + \frac{1}{2^{9(s+1)}(2^{s+1}-1)}\right) \\ &+ \frac{e^{3/K} K^{\ell} \ell!}{2} \left(\sum_{k\geqslant 1} \prod_{s=1}^{k} \left(\rho + \frac{c_1}{512 \cdot 2^{(s+1)}} + \frac{1}{2^{9(s+1)}(2^{s+1}-1)}\right) \right)^2 \\ &= 0.141422204463 \cdots K^{\ell} \ell!. \end{aligned} \tag{21}$$

Combining (19),(20), and (21) gives:

$$d_{\ell,1} + d_{\ell,2} + d_{\ell,3} \leqslant 0.795685442972 \cdots K^{\ell} \ell!.$$

Finally, for $\ell \geqslant 10$,

$$\begin{aligned} d_{\ell} &\leqslant \frac{d_{\ell,1} + d_{\ell,2} + d_{\ell,3}}{g_{10}(\rho)} - \frac{g'_{10}(\rho)}{g_{10}(\rho)} c_{\ell} \\ &\leqslant 0.869090272578 \cdots K^{\ell} \ell! \end{aligned}$$

since $-g'_{\ell}(\rho)/g_{\ell}(\rho)$ decreases to 0 and $g_{10}(\rho) = 0.986940779096 \cdots$ and $-g'_{10}(\rho)/g_{10}(\rho) = 0.062876303286 \cdots$. This proves the claim. □

A.2 The boundary of a phylogenetic network

Recall from Section 3.2 that, to extend the definition of B_2 to infinite phylogenetic networks, we need a notion of "boundary" of a network that includes not only the vertices in which the directed random walk X can get trapped (i.e. the leaves), but also all of the distinct ways in which it can escape to infinity (i.e. the ends, see Definition 3.3). This motivated Definition 3.4 of the boundary ∂G of a phylogenetic network G as $\partial G = \mathcal{L} \cup \mathcal{E}$, where $\mathcal{L}$ and $\mathcal{E}$ denote the set of leaves and the set of ends of G, respectively.

In this appendix, we show that ∂G can be embedded in a suitable compact metric space $(\mathcal{K},\, d_{\mathcal{K}})$, and that the escape point X_∞ of the directed random walk X is a well-defined random variable in $\mathcal{K}$. We also discuss some topological properties of the boundary ∂G, in particular its connection with the Martin boundary.

Let us start by introducing some notation. We write $u \preccurlyeq v$ to indicate that u is an ancestor of v, i.e. that there exists a finite directed path from u to v in G. For any vertex $v \in G$, we denote by

$$\bar{v} \;=\; \{u \in G : u \preccurlyeq v\}$$

the set of ancestors of v. Finally, for any set of vertices $S \subset G$, we use the short notation $[S]_k$ for the set $S \cap [G]_k$ of vertices of S at height at most k in G.

Definition A.3. Let G be a phylogenetic network. Set

$$\mathcal{K} \;=\; \{S \subset G : \forall u \in S,\, \bar{u} \subset S\}\,,$$

and equip $\mathcal{K}$ with the distance $d_{\mathcal{K}}$ defined by

$$d_{\mathcal{K}}(S, S') \;=\; \sup\Big\{2^{-n} : n \in \mathbb{N},\; [S]_n = [S']_n\Big\},$$

with the convention $\sup \emptyset = 0$. ◇

It is readily checked that $(\mathcal{K}, d_{\mathcal{K}})$ is a compact metric space (using a diagonal argument and the local finiteness of G to show compactness).

Because the map $v \in G \mapsto \bar{v} \in \mathcal{K}$ is injective, the vertices of G can be seen as points of $\mathcal{K}$. The next proposition shows that the ends of G can also be embedded in $\mathcal{K}$; more precisely, they are points of $\mathrm{cl}(G)$, the topological closure of G in $\mathcal{K}$.

Proposition A.4. *Let G be a phylogenetic network, and let $\mathcal{E}$ denote its end set. For each $x \in \mathcal{E}$, define*

$$\bar{x} \;=\; \{v \in G : \exists r \in x \text{ such that } v \in r\},$$

where $v \in r$ means that the ray r goes through the vertex v. Then, the map

$$x \in \mathcal{E} \;\longmapsto\; \bar{x} \in \mathcal{K}$$

is injective. Furthermore, for any ray $r = (v_n)_{n \geqslant 0}$, we have $\bar{v}_n \to \bar{x}$ in $(\mathcal{K}, d_{\mathcal{K}})$, where x is the end associated with r.

Proof. First, let us show that $\bar{x} \in \mathcal{K}$ for any $x \in \mathcal{E}$. Fix $x \in \mathcal{E}$, $v \in \bar{x}$ and $u \preccurlyeq v$. We must show that $u \in \bar{x}$, i.e. that x contains a ray that goes through u. Let $r \in x$

be such that $v \in r$. Because G is rooted and $u \preccurlyeq v$, there exists a path that goes from the root of G to u, and then from u to v. Let r' be the ray that starts with this path and then continues along r. Since $r' \# r$, we have $r' \in x$, concluding the proof that $\bar{x} \in \mathcal{K}$.

Let us now show that $x \mapsto \bar{x}$ is injective. Consider $x, y \in \mathcal{E}$ such that $\bar{x} = \bar{y}$, and fix two rays $r = (v_i)_{i \geqslant 0} \in x$, and $r' = (v'_i)_{i \geqslant 0} \in y$. To show that $x = y$, we must show that $r \# r'$, and for this it suffices to exhibit a ray that intersects r and r' infinitely many times. The key is to show that for all $i \geqslant 0$, there exists $j \geqslant 0$ such that $v_i \preccurlyeq v'_j$ – indeed, by symmetry we then get that there exists $k \geqslant 0$ such that $v'_j \preccurlyeq v_k$, etc; and by a straightforward concatenation procedure we can build a ray that goes to and fro between vertices of r and vertices of r'. By assumption, $\bar{x} = \bar{y}$ and thus $v_i \in \bar{y}$. As a result, there exists a ray $r'' \# r'$ such that $v_i \in r''$, and since r'' intersects r' infinitely many times, there exists $u \in r''$, $u \succcurlyeq v_i$ such that $u \in r'$, concluding the proof of the injectivity of $x \mapsto \bar{x}$.

Finally, consider any ray $r = (v_n)_{n \geqslant 0}$, letting x denote the corresponding end. Let us show that $\bar{v}_n \to \bar{x}$ as $n \to \infty$, i.e. let us fix some $k \geqslant 0$ and show that $[\bar{x}]_k = [\bar{v}_i]_k$ for all i large enough. Let $v \in [\bar{x}]_k$, and pick $r_v \in x$ such that $v \in r_v$. Because $r_v \# r$, there exists $i_v \in \mathbb{N}$ such that $v \preccurlyeq v_{i_v}$. Since $[G]_k$ is finite, we can define

$$N_k \;=\; \max\{i_v : v \in [\bar{x}]_k\}\,.$$

Thus, $[\bar{x}]_k \subset \bar{v}_i$ for all $i \geqslant N_k$. Moreover, because $\bar{x} \in \mathcal{K}$ we also have $\bar{v}_i \subset \bar{x}$ for all $i \geqslant 0$. As a result, $[\bar{x}]_k = [\bar{v}_i]_k$ for all $i \geqslant N_k$, concluding the proof. ☐

Proposition A.4 shows that the vertices and the ends of G can be embedded in $\mathcal{K}$ (in the rest of this section, we thus identify $v \in G$ with $\bar{v} \in \mathcal{K}$ and simply drop the notation $\bar{v}$), and it immediately implies the following corollary.

Corollary A.5. *Let G be a phylogenetic network and let $(X_t)_{t \geqslant 0}$ be the directed random walk on G. Then, X_t converges almost surely in $(\mathcal{K}, d_{\mathcal{K}})$ to a random variable $X_\infty \in \partial G$.*

Remark A.6. Note that there is no connection between the boundary ∂G and the topological boundary of G in $\mathcal{K}$, that is, $\mathrm{cl}(G) \setminus \mathrm{int}(G)$. Indeed:

(i) Because the vertices of G are isolated points in $\mathcal{K}$, they all belong to the interior of G, i.e. G is open in $\mathcal{K}$ – whereas leaves belong to ∂G.

(ii) Although $\mathcal{E} \subset \mathrm{cl}(G)$, not all limit points are ends: as Figure 9 shows, limit points can correspond to union of ends. ◇

We close this appendix by briefly discussing the connection with a standard notion of boundary for transient random walks: the Martin boundary. Let us start by recalling its definition in our setting; we refer the reader to [17, Chapter 10] for a more general definition.

Let $\mathbb{P}_u(\,\cdot\,)$ denote probabilities conditional on the directed random walk X being started from u, instead of from the root. The *Martin kernel of* X is defined as

$$M(u, v) \;=\; \frac{\mathbb{P}_u(v \in X)}{\mathbb{P}(v \in X)}\,.$$

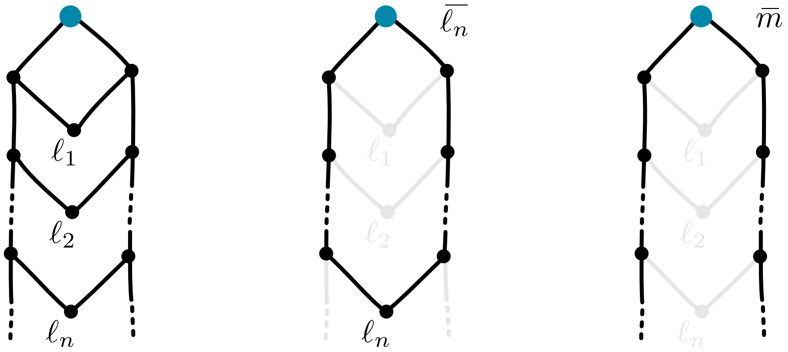

Figure 9: As previously, the root is highlighted in blue and the edges are pointing downwards. In this example, the sequence (ℓ_n) has a limit $m \in \mathcal{K} \setminus \partial G$. Note that this limit corresponds to the union of two ends: in fact, although we do not detail this here, it is not too hard to show that the limit in $\mathcal{K}$ of a sequence of points of ∂G is always either a leaf or a union of ends.

Note that the functions $M(u, \cdot\,)$ are bounded, since

$$\frac{\mathbb{P}_u(v \in X)}{\mathbb{P}(v \in X)} = \frac{\mathbb{P}(u \in X \mid v \in X)\, \mathbb{1}_{\{u \preccurlyeq v\}}}{\mathbb{P}(u \in X)} \leqslant \frac{1}{\mathbb{P}(u \in X)}\,.$$

As a result, there is a canonical compactification of G on which all functions $M(u, \cdot\,)$ extend continuously (see e.g. [28, Theorem 7.3]). This compactification, which we denote by $\widehat{G}$, is known as the Martin compactification of G. The *Martin boundary of* G is then defined as $\widehat{G} \setminus G$.

Equivalently, a sequence (v_n) of vertices of G converges in $\widehat{G}$ if and only if $K(u, v_n)$ converges for all $u \in G$. The Martin boundary consists of all limit points of sequences (v_n) that escape to infinity (in the sense that $h(v_n) \to \infty$, where $h(v)$ denotes the height of v).

Although the Martin boundary and our boundary ∂G are both designed to capture the asymptotic behavior of X (and both make it possible to define an almost-sure limit for X_t), they are distinct notions and there does not seem to be a simple relationship between them; in particular, one is not finer than the other.

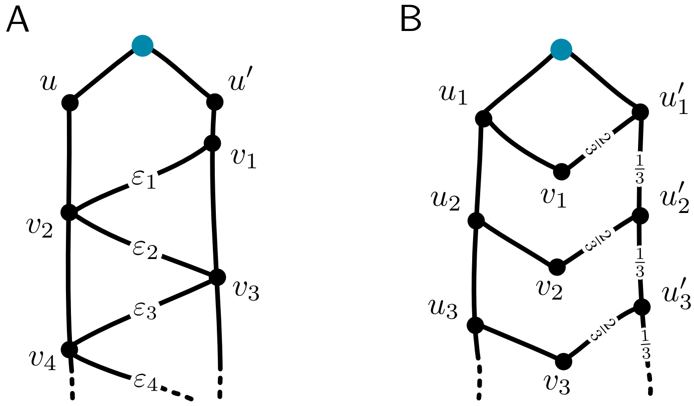

Figure 10: Two phylogenetic networks for which the Martin boundary differs from $\mathcal{K}$. Again, the root is highlighted in blue and the edges are pointing downwards. Provided that the weights of the edges going out of a vertex are rational probabilities, by replacing that vertex by a network as in Section 3.3, we can assume that the probability that the directed random walk follows an edge is equal to its weight.

For instance, the phylogenetic network G given in Figure 10.A provides an example of a sequence of vertices that converges in $\mathcal{K}$ to an end $x \in \partial G$, but that does not converge in $\widehat{G}$: indeed, assume that the weights (ε_n) given in Figure 10.A satisfy

$$\delta \;:=\; \sum_{n \geqslant 1} \varepsilon_n \;\leqslant\; \frac{1}{4}\,.$$

Let A be the event that the directed random walk X started from the root of G goes through one of the central edges $(v_1 \to v_2), (v_2 \to v_3), \ldots$, that is

$$A \;=\; \bigcup_{k \geqslant 1} \Big\{ v_k \in X \text{ and } v_{k+1} \in X \Big\}\,.$$

Using union bounds, we see that

$$\max\Big\{\mathbb{P}(A),\; \mathbb{P}(A \,|\, u \in X),\; \mathbb{P}(A \,|\, u' \in X)\Big\} \;\leqslant\; \delta\,.$$

If X goes to u and then never crosses to the right through one of the central edges, then it goes through v_{2n}. As a result, letting A^{c} denote the complement of A,

$$\mathbb{P}\Big(u \in X, v_{2n} \in X\Big) \;\geqslant\; \mathbb{P}\Big(\{u \in X\} \cap A^{\mathsf{c}}\Big) \;\geqslant\; \frac{1}{2}(1-\delta)\,.$$

Conversely, if X goes to u and then visits v_{2n+1}, then it must have crossed to the right using one of the central edges – so that

$$\mathbb{P}\Big(u \in X, v_{2n+1} \in X\Big) \;\leqslant\; \mathbb{P}\Big(\{u \in X\} \cap A\Big) \;\leqslant\; \frac{1}{2}\,\delta\,.$$

Similarly, we get $\mathbb{P}(u' \in X, v_{2n} \in X) \leqslant \frac{1}{2}\delta$ and $\mathbb{P}(u' \in X, v_{2n+1} \in X) \geqslant \frac{1}{2}(1-\delta)$. As a result,

$$\begin{cases} \mathbb{P}\Big(u \in X \,\Big|\, v_{2n} \in X\Big) \;\geqslant\; \frac{1}{2}(1-\delta) \;\geqslant\; \frac{3}{8} \\ \mathbb{P}\Big(u \in X \,\Big|\, v_{2n+1} \in X\Big) \;\leqslant\; \frac{\delta}{1-\delta} \;\leqslant\; \frac{1}{3}\,, \end{cases}$$

from which it follows that (v_n) does not converge in $\widehat{G}$ – whereas it does in $\mathcal{K}$, since ∂G consists of a single end.

Conversely, the phylogenetic network G' given in Figure 10.B provides an example of a sequence that converges in the Martin boundary $\widehat{G}'$, but not in $\mathcal{K}'$: indeed, it is readily computed that

$$\mathbb{P}\Big(u_1 \in X \,\Big|\, v_n \in X\Big) \;=\; \frac{(\frac{1}{2})^{n+1}}{(\frac{1}{2})^{n+1} + (\frac{1}{3})^{n}} \;\xrightarrow[n\to\infty]{}\; 1,$$

from which it follows that the sequence $(u_1, v_1, u_2, v_2, \ldots)$ converges in $\widehat{G}'$. Yet this sequence does not converge in $\mathcal{K}$, because the subsequences (u_n) and (v_n) have different limits.

A.3 Properties of the entropy

To make this article self-contained, in this appendix we recall the definition of the entropy of a probability measure on a general (i.e. not necessarily countable) space, as well as some of its main properties. The results are presented mostly without proofs, and we refer the reader to [21, Chapter 2] for a detailed treatment.

Let us start by recalling the definition of the entropy of a countable partition.

Definition A.7. Let $(\Omega, \mathscr{A}, \mu)$ be a probability space, and let π be a countable measurable partition of Ω. Then, the *entropy of π with respect to μ is*

$$H_\mu(\pi) \;=\; -\sum_{A\in\pi} \mu(A)\log_2 \mu(A)\,. \qquad \diamond$$

A key property of the entropy, which we refer to as its *monotonicity*, is given by the next proposition. Recall that a partition π is said to be *finer* than a partition π' if for all $B \in \pi$ there exists $B' \in \pi'$ such that $B \subset B'$. We write $\pi \preccurlyeq \pi'$ to indicate that the partition π is finer than π'.

Proposition A.8 (Monotonicity of the entropy)**.** *Let π and π' be two [countable] measurable partitions. If $\pi' \preccurlyeq \pi$, then $H_\mu(\pi') \geqslant H_\mu(\pi)$.*

In particular, Proposition A.8 justifies – and is trivially preserved by – the following extension of Definition A.7 to uncountable partitions.

Definition A.9. Let $(\Omega, \mathscr{A}, \mu)$ be a probability space, and let π be a measurable partition of Ω. The *entropy of π with respect to μ* is

$$H_\mu(\pi) \;=\; \sup\Big\{H_\mu(\pi') : \pi' \text{ countable measurable partition of } \Omega \text{ s.t. } \pi \preccurlyeq \pi'\Big\}\,.$$

The *entropy of the probability distribution* μ is then defined as

$$H(\mu) \;=\; \sup\Big\{H_\mu(\pi) : \pi \text{ measurable partition of } \Omega\Big\}\,. \qquad \diamond$$

The monotonicity of the entropy entails that, letting σ denote the partition into singletons, $H(\mu) = H_\mu(\sigma)$. In particular, if Ω is countable then we recover the familiar definition of Shannon entropy:

$$H(\mu) \;=\; -\sum_{i\in\Omega} \mu(\{i\})\log_2 \mu(\{i\})\,.$$

Moreover, in the case of a standard probability space (that is, a Polish space equipped with its Borel σ-field), Definition A.9 takes a simple form – in fact, this simple form is sometimes used to define the entropy, see e.g. [21, Definition 2.14]. Before stating it, let us point out an elementary but useful fact. Since the proof only involves straightforward calculations, we omit it.

Proposition A.10. *Let $(\Omega, \mathscr{A}, \mu)$ be a probability space, and let π be a measurable partition of Ω. Assume that π' is obtained from π by fragmenting one of its block B such that $\mu(B) > 0$, and let π_B denote the corresponding partition of B. Then,*

$$H_\mu(\pi') \;=\; H_\mu(\pi) \;+\; \mu(B)\, H_{\mu_B}(\pi_B)\,,$$

where μ_B denotes the conditional probability distribution induced on B by μ.

Proposition A.11. *Let $(\Omega, \mathscr{A}, \mu)$ be a standard probability space. If the support of μ is uncountable, then $H(\mu) = +\infty$.*

Proof. Let A denote the set of atoms of μ, noting that by assumption $\mu(A^{\mathsf{c}}) > 0$. Let σ denote the partition of Ω into singletons, and let

$$\sigma_A \;=\; \{\{i\} : i \in A\} \cup \{A^{\mathsf{c}}\}\,.$$

By Proposition A.10, we have

$$H(\mu) \;=\; H_\mu(\sigma_A) \;+\; \mu(A^{\mathsf{c}})\,H(\mu_{A^{\mathsf{c}}})\,.$$

Thus, to finish the proof it suffices to show that $H(\mu_{A^{\mathsf{c}}}) = +\infty$.

By the Borel isomorphism theorem, there exists a Borel isomorphism ψ that maps A^{c} to an interval of $I \subset \mathbb{R}$. Let ν be the pushforward of $\mu_{A^{\mathsf{c}}}$ by ψ. Since ν has no atoms, its cumulative distribution function is continuous on I. Therefore, for any $n \geqslant 1$, it is possible to partition I into n intervals $I_1, \dots, I_n$ such that $\nu(I_n) = 1/n$. Let then $\pi_n = \{\psi^{-1}(I_k) : k = 1, \dots, n\}$. Thus, π_n is a finite measurable partition of A^{c} such that $\mu_{A^{\mathsf{c}}}(B) = 1/n$ for all $B \in \pi_n$. It follows that

$$H(\mu_{A^{\mathsf{c}}}) \;\geqslant\; H_{\mu_{A^{\mathsf{c}}}}(\pi_n) \;=\; \log_2(n)\,.$$

Since n can be arbitrarily large, this concludes the proof. □

Finally, we close this section by a useful lemma for refining sequences of partitions. For this, we first introduce some vocabulary and notations. Given a partition π, we denote by:

- $\pi[x]$ the block of π that contains x;
- $\sim_\pi$ the equivalence relation associated with π.

Definition A.12. We say that a sequence (π_n) of partitions of a set E *converges to the partition* π if, for all $x, y \in E$, there exists N such that, for all $n \geqslant N$,

$$x \sim_{\pi_n} y \iff x \sim_\pi y\,.$$ ◇

Remark A.13. Equivalently, (π_n) converges to π if and only if $\lim_n \pi_n[x] = \pi[x]$ for every $x \in E$, i.e. if and only if $\mathbb{1}_{\pi_n[x]}$ converges pointwise to $\mathbb{1}_{\pi_n[x]}$ for all $x \in E$. ◇

Lemma A.14. *Let (π_n) be a sequence of measurable partitions on a standard probability space $(\Omega, \mathscr{A}, \mu)$. If (π_n) converges to a partition π, such that $\pi \preccurlyeq \pi_n$ for all $n \geqslant 0$, then π is measurable and*

$$H_\mu(\pi_n) \xrightarrow[n\to\infty]{} H_\mu(\pi).$$

In particular, if π is the partition into singletons, the limit is $H(\mu)$.

Proof. The fact that π is measurable can be deduced from Remark A.13. This remark is also the key to the rest of the proof. Indeed, since $\pi_n[x] \to \pi[x]$ for all $x \in \Omega$,

$$\mu(\pi_n[x]) \xrightarrow[n\to\infty]{} \mu(\pi[x]), \qquad \forall x \in \Omega. \tag{22}$$

Now, for any measurable partition σ, define the following measurable function:

$$f^\sigma \colon\; x \in \Omega \longmapsto \begin{cases} -\log_2\big(\mu(\sigma[x])\big) & \text{if } \mu(\sigma[x]) > 0, \\ \infty & \text{otherwise.} \end{cases}$$

By Equation (22), $f^{\pi_n} \to f^\pi$ pointwise. Noting that, in view of Proposition A.11, $H_\mu(\sigma) = \int f^\sigma\, d\mu$, we deduce from Fatou's lemma that

$$\liminf_{n\to\infty} H_\mu(\pi_n) \;\geqslant\; H_\mu(\pi).$$

This concludes the proof, since, by Proposition A.8, $H_\mu(\pi_n) \leqslant H_\mu(\pi)$ for all n. □

A.4 Properties of the B_2 index

In this Appendix, we list and prove miscellaneous properties of the B_2 index. Let us start with something elementary, but fundamental – namely, the fact that the grafting property holds for infinite phylogenetic networks.

Proposition A.15. *Let G_1 and G_2 be two phylogenetic networks, and let G be the phylogenetic network obtained by identifying a leaf $\ell \in G_1$ with the root of G_2. Then,*

$$B_2(G) \;=\; B_2(G_1) \;+\; p_\ell \, B_2(G_2)\,,$$

where p_ℓ denotes the probability that the directed random walk on G_1 ends in ℓ.

Since the proof of Proposition A.15 is a straightforward application of the analogous Proposition A.10 for the entropy, we do not detail it.

Next, let us prove the measurability of B_2. This technicality is needed to ensure that the B_2 index of a random phylogenetic network is a well-defined random variable.

Proposition A.16. *Let $\mathbb{G}$ denote the space of phylogenetic networks, equipped with the local topology and its Borel σ-field. The function $B_2 \colon \mathbb{G} \to \mathbb{R}$ is measurable.*

Proof. We will show that B_2 can be expressed as a pointwise limit of continuous functions from $\mathbb{G}$ to $\mathbb{R}$.

Let G be a fixed phylogenetic network, and let $X = (X_n)_{n \geqslant 0}$ be the directed random walk on G, started from the root. Note that X is the only source of randomness, and that, for any $n \geqslant 0$, the distribution of X_n depends only on $[G]_n$. Therefore, the functions

$$B_2^{k,n} \colon G \;\longmapsto\; -\sum_{A \subset [G]_k} p_A^n \log_2(p_A^n), \quad \text{where } p_A^n = \mathbb{P}\Big([\bar{X}_n]_k = A\Big),$$

are continuous for the local topology (recall that $\bar{X}_n$ denotes the set of ancestors of X_n in G, and that we use the short notation $[\bar{X}_n]_k = \bar{X}_n \cap [G]_k$). Furthermore, we gave in Definition A.3 a compactification $\mathcal{K}$ of G in which X_n converges almost surely to X_∞ and in which, by definition, the functions $x \mapsto [\bar{x}]_k$ are continuous, for all k. Therefore, for any $k \geqslant 1$ and $A \subset [G]_k$, by dominated convergence,

$$p_A^n \;=\; \mathbb{P}\Big([\bar{X}_n]_k = A\Big) \;\xrightarrow[n\to\infty]{}\; \mathbb{P}\Big([\bar{X}_\infty]_k = A\Big) \;=:\; p_A\,.$$

This implies that

$$B_2^{k,n}(G) \;\xrightarrow[n\to\infty]{}\; B_2^{k}(G) \;:=\; -\sum_{A \subset [G]_k} p_A \log_2(p_A) \;=\; H_\mu(\pi_k),$$

where μ is the distribution of X_∞ on $\mathcal{K}$, and π_k is the partition defined by

$$x \sim_{\pi_k} y \quad\Longleftrightarrow\quad [\bar{x}]_k = [\bar{y}]_k\,.$$

Because the map $x \mapsto \bar{x}$ is injective (see Proposition A.4), as $k \to \infty$ the partition π_k tends to the partition into singletons. Therefore, by Lemma A.14 and definition of B_2,

$$B_2^k(G) = H_\mu(\pi_k) \;\xrightarrow[k\to\infty]{}\; H(\mu) = B_2(G),$$

concluding the proof. ☐

We now provide tools to study the continuity of B_2. We focus on easy-to-check sufficient conditions that are likely to hold for biologically relevant models of phylogenetic networks.

Proposition A.17. *Let (G_n) be a sequence of phylogenetic networks converging locally to G. Assume that the directed random walk on G ends in a leaf almost surely. Then,*

$$\liminf_{n\to\infty} B_2(G_n) \;\geqslant\; B_2(G).$$

Proof. Let $\mathcal{L}$ be the leaf set of G, and let $\mathcal{L}_k = \{\ell \in \mathcal{L} : h(\ell) \leqslant k\}$ be the set of leaves at height at most k. As always, write X for the directed random walk on G, and X_∞ for its limit on the boundary of G. Finally, let $p_\ell = \mathbb{P}(X_\infty = \ell)$.

Since $\mathbb{P}(X_\infty \in \mathcal{L}) = 1$, we have $B_2(G) = -\sum_{\ell\in\mathcal{L}} p_\ell \log_2 p_\ell$; and since $\mathcal{L} = \lim\!\uparrow_k \mathcal{L}_k$, it follows that

$$-\sum_{\ell\in\mathcal{L}_k} p_\ell \log_2 p_\ell \;\xrightarrow[k\to\infty]{}\; B_2(G)\,.$$

Now, fix some b such that $0 < b < B_2(G)$, and let k be such that the sum in the display above be greater than b. Fix n_0 such that $[G_n]_k = [G]_k$ for all $n \geqslant n_0$. Then, for all $n \geqslant n_0$,

$$B_2(G_n) \;\geqslant\; -\sum_{\ell\in\mathcal{L}_k} p_\ell \log_2 p_\ell \;>\; b.$$

This shows that $\liminf_n B_2(G_n) \geqslant b$. Letting $b \to B_2(G)$ concludes the proof. ☐

Proposition A.18. *Let (G_n) be a sequence of* finite *phylogenetic networks that converge locally to G. Assume that, for a sequence (k_n) satisfying $[G_n]_{k_n} = [G]_{k_n}$, we have*

$$\mathbb{P}\Big(h(X_\infty) > k_n\Big)\,\log|G_n| \;\xrightarrow[n\to\infty]{}\; 0\,,$$

where $h(X_\infty)$ is the height of the limit of the directed random walk on G, and $|G_n|$ denotes the number of leaves of G_n. Then,

$$\lim_{n\to\infty} B_2(G_n) \;=\; B_2(G).$$

Proof. Let us write $q_n = \mathbb{P}(h(X_\infty) > k_n)$. Since $\log|G_n|$ cannot tend to 0 (except in the trivial case where G is a finite DAG with a single leaf), we must have $q_n \to 0$. This means that $\mathbb{P}(X_\infty \in \mathcal{L}) = 1$, where $\mathcal{L}$ denotes the leaf set of G. As a result, by Proposition A.17 it suffices to show

$$\limsup_{n\to\infty} B_2(G_n) \;\leqslant\; B_2(G) \;=\; -\sum_{\ell\in\mathcal{L}} p_\ell \log_2 p_\ell\,,$$

where $p_\ell = \mathbb{P}(X_\infty = \ell)$. Let us write $\mathcal{L}^n_{k_n}$ (resp. $\mathcal{L}^n_{>k_n}$) for the set of leaves in G_n at height at most k_n (resp. at least $k_n + 1$). For any leaf $\ell \in G_n$, write $p^{(n)}_\ell$ for the probability that the directed random walk on G_n ends in ℓ. Notice that, since $[G_n]_{k_n} = [G]_{k_n}$,

$$-\sum_{\ell\in\mathcal{L}^n_{k_n}} p^{(n)}_\ell \log_2 p^{(n)}_\ell \;=\; -\sum_{\ell\in\mathcal{L}_{k_n}} p_\ell \log_2 p_\ell\,,$$

and that, since $(p^{(n)}_\ell/q_n : \ell \in \mathcal{L}^n_{>k_n})$ is a probability distribution, Jensen's inequality gives

$$-\frac{1}{q_n}\sum_{\ell\in\mathcal{L}^n_{>k_n}} p^{(n)}_\ell \log_2 p^{(n)}_\ell \;\leqslant\; \log_2\Bigg(\sum_{\ell\in\mathcal{L}^n_{>k_n}} \frac{p^{(n)}_\ell}{q_n}\cdot\frac{1}{p^{(n)}_\ell}\Bigg) \;\leqslant\; \log_2\left(\frac{|G_n|}{q_n}\right).$$

It follows that

$$B_2(G_n) \;\leqslant\; -\sum_{\ell\in\mathcal{L}_{k_n}} p_\ell \log_2 p_\ell \;+\; q_n \log_2\left(\frac{|G_n|}{q_n}\right),$$

where the first term on the right-hand side tends to $B_2(G)$, while the second one tends to 0 because $q_n \to 0$ and $q_n \log_2 |G_n| \to 0$ by hypothesis. □

Finally, we close this section by providing a useful tool to compute the B_2 index of a fixed infinite network G. This essentially consists in chopping the network at its cut-vertices, and using the grafting property – but there is a small subtlety (namely, showing that the resulting nondecreasing sequence of B_2 indices, which is trivially bounded above by the $B_2(G)$, actually reaches it in the limit). In order to give a clean statement, let us introduce some vocabulary.

Definition A.19. A *stub* of a phylogenetic network G is any phylogenetic network $S \subset G$ that can be obtained by removing all out-going edges from a set of cut-vertices of G and keeping the connected component containing the root. We write $S \sqsubset G$ to indicate that S is a stub of G. A *stub sequence* of G is an increasing sequence of stubs $S_n \sqsubset S_{n+1} \sqsubset G$ such that $\lim_n S_n = G$. ◇

Proposition A.20. *Let (S_n) be a stub sequence of G. Then,*

$$B_2(S_n) \xrightarrow[n\to\infty]{} B_2(G)\,.$$

Proof. For any stub $S \sqsubset G$, let π_S be the partition of ∂G defined by the following equivalence relation:

$$x \sim_S y \iff x = y \text{ or } \bar{x}\cap\bar{y} \text{ contains a leaf of } S.$$

To see that this is indeed an equivalence relation, recall that the leaves of S are either leaves or cut-vertices in G, and note that for all $x \in \partial G$, the set $\bar{x}$ contains at most one leaf of S. The transitivity of $\sim_S$ then follows readily.

Note that the partitions π_{S_n} are refining along the stub sequence (S_n), and that they converge simply (see Definition A.12) to π_G, the partition into singletons. Indeed, pick $x \neq y$ and assume that $x \sim_{S_n} y$ for all n. For each n, let then v_n be the unique leaf of S_n in $\bar{x}\cap\bar{y}$. Then:

- If $\{v_n : n \geqslant 1\}$ is finite, then there exists ℓ such that $v_n = \ell$ for all n large enough. By the local convergence of S_n to G, this vertex ℓ is a leaf of G, and therefore $x = y = \ell$, yielding a contradiction.
- If $\{v_n : n \geqslant 1\}$ is infinite, then $v_1, v_2, \ldots$ are cut-vertices of G and they lie on a ray $r \subset \bar{x}\cap\bar{y}$. This entails $x = y$: indeed, let $r_x \in x$ and $r_y \in y$. For all n, since $v_n \in \bar{x}$ and v_n is a cut-vertex of G, $v_n \in r_x$. As a result, r_x intersects r infinitely many times. Similarly, r_y intersects r infinitely many times. Thus, we have $r_x \;\#\!\!\#\; r_y$ and therefore $x = y$ – again yielding a contradiction.

Finally, note that $B_2(S_n) = H_\mu(\pi_{S_n})$, where μ is the distribution of X_∞, and that, by definition, $B_2(G) = H_\mu(\pi_G)$. Thus, the proposition follows from Lemma A.14. □

A.5 Total variation bound for the convergence of Galton–Watson trees

In this appendix, we provide a total variation bound to quantify the speed of convergence of conditioned Galton–Watson trees to Kesten's tree. Let us start by setting / recalling some notation.

In the remainder of this appendix, ξ will denote a random variable taking values in $\mathbb{N} = \{0, 1, 2, \ldots\}$, with mean 1 and such that $\mathbb{P}(\xi = 0) > 0$. Likewise, $T \sim \mathrm{GW}(\xi)$ will be a Galton–Watson tree with offspring distribution ξ, which we view as an ordered tree. Recall that, for any tree $\mathbf{t}$, we denote the number of leaves of $\mathbf{t}$ by $|\mathbf{t}|$. As in Section 3.3, for any n such that $\mathbb{P}(|T| = n) > 0$, we write T_n for a random tree distributed as T conditioned to have n leaves. Note that, whenever we state something about T_n, we implicitly assume that n is taken so that $\mathbb{P}(|T| = n) > 0$. Finally, we denote by T_* the Kesten tree associated to T, i.e. the local limit of T_n as $n \to \infty$ (see Definition 3.6), and we write $\hat{\xi}$ for a random variable such that $\mathbb{P}(\hat{\xi} = n) = n\,\mathbb{P}(\xi = n)$ for all $n \geqslant 1$ (thus, $\hat{\xi}$ is distributed as the number of children of the vertices that lie on the spine of T_*).

Let us now introduce a truncation of T_*: for each $k \geqslant 0$, let v_k be the k-th vertex on the spine of T_*, with v_0 being the root, and let T_*^k be the leaf-pointed rooted tree obtained from T_* by removing all descendants of v_k (other than v_k itself) and letting v_k be the distinguished leaf. The distribution of T_*^k is easily computed: let $\mathbf{t}$ denote a finite ordered tree with a distinguished leaf v_k at height k, and write $\mathbf{v} = (v_0, v_1, \ldots, v_k)$ for the path from the root of $\mathbf{t}$ to its distinguished leaf v_k. Then, by construction of T_*,

$$\begin{aligned}
\mathbb{P}\Big(T_*^k = \mathbf{t}\Big) &= \left(\prod_{i=0}^{k-1} \mathbb{P}\Big(\hat{\xi} = d_{v_i}\Big) \cdot \frac{1}{d_{v_i}}\right)\left(\prod_{v \in \mathbf{t} \setminus \mathbf{v}} \mathbb{P}(\xi = d_v)\right) \\
&= \prod_{v \in \mathbf{t} \setminus \{v_k\}} \mathbb{P}(\xi = d_v) \\
&= \frac{\mathbb{P}(T = \mathbf{t})}{\mathbb{P}(\xi = 0)},
\end{aligned}$$

where d_v denotes the number of children of vertex v and, by a slight abuse of notation, we write $\{T = \mathbf{t}\}$ for the event that the Galton–Watson tree T is equal to the non-pointed version of the tree $\mathbf{t}$. Moreover, note that $\mathbb{P}(T = \mathbf{t})/\mathbb{P}(\xi = 0)$ is also the probability that T "starts" with the leaf-pointed tree $\mathbf{t}$, i.e. that T can be obtained by grafting a tree T' on the distinguished leaf of $\mathbf{t}$. Writing $\{\mathbf{t} \subset T\}$ for this event, we therefore have

$$\mathbb{P}\Big(T_*^k = \mathbf{t}\Big) = \mathbb{P}(\mathbf{t} \subset T).$$

Also note that, conditional on $\{\mathbf{t} \subset T\}$, the subtree of T descending from the distinguished leaf of $\mathbf{t}$ is a $\mathrm{GW}(\xi)$ tree.

Next, let us build a pointed tree T_n^k from T_n, as we did for T_*^k. Conditional on T_n, let u_n be chosen uniformly at random among the n leaves of T_n. If u_n is at height at least k, then let $v_{n,k}$ be the vertex at distance k from the root on the path to u_n, and define T_n^k to be the tree obtained from T_n by removing all vertices descending from $v_{n,k}$ (other than $v_{n,k}$ itself) and letting $v_{n,k}$ be the distinguished vertex. On the event that u_n is at height less than k, the definition of T_n^k is irrelevant; we define it to be the tree reduced to a single node. From this construction, note that for any $\mathbf{t}$ finite ordered tree with a distinguished leaf at height k,

$$\begin{aligned}\mathbb{P}\Big(T_n^k = \mathbf{t}\Big) &= \frac{n - |\mathbf{t}| + 1}{n} \cdot \frac{\mathbb{P}(\mathbf{t} \subset T,\, |T| = n)}{\mathbb{P}(|T| = n)} \\ &= \frac{n - |\mathbf{t}| + 1}{n} \cdot \mathbb{P}(\mathbf{t} \subset T) \cdot \frac{\mathbb{P}(|T| = n - |\mathbf{t}| + 1)}{\mathbb{P}(|T| = n)} \\ &= \mathbb{P}\Big(T_*^k = \mathbf{t}\Big) \cdot \frac{f(n - |\mathbf{t}| + 1)}{f(n)}, \end{aligned} \tag{23}$$

where $f(n) := n\,\mathbb{P}(|T| = n)$.

With these definitions, we are ready to state the main result of this appendix.

Proposition A.21. *If $\mathbb{E}(\xi^3) < \infty$, then for any sequence of integers $(k_n)_{n \geqslant 0}$ such that $k_n = o(\sqrt{n})$, we have*

$$d_{\mathrm{TV}}\Big(T_n^{k_n},\, T_*^{k_n}\Big) \;=\; \Theta\Big(\frac{k_n}{\sqrt{n}}\Big).$$

Remark A.22. The fact that the total variation distance between $T_n^{k_n}$ and $T_*^{k_n}$ tends to 0 whenever $k_n = o(\sqrt{n})$ already appears, in slightly different forms, in the literature: a similar result, but where T_n is conditioned on its number of vertices, can be found in Kersting [18, Theorem 5] – without any moment assumption besides $\mathbb{E}(\xi) = 1$, and for any sequence $k_n = o(n/a_n)$ with a_n being any sequence such that $\sum_{i=1}^n \xi_i / a_n$ converges in distribution (where ξ_i are i.i.d. replicates of ξ). Stufler [25, Theorem 5.2] also states a closely related result for fringe subtrees, under a finite variance assumption. What is new here is the explicit speed of convergence $k_n/\sqrt{n}$, which we need to show the convergence of all moments of $B_2(G_n)$ to those of $B_2(G_*)$ in the proof of Theorem 3.7.

The third moment assumption is needed in our proof to get a Berry–Esseen type estimate, but this may be a superfluous assumption. It would be interesting to see whether similar bounds can be obtained in the critical case without any moment assumption (similar to Kersting), but this goes beyond the scope of this paper.

Finally, note that the proof that we develop would be essentially the same if T_n were conditioned on the number of vertices, or if we looked at fringe subtrees. ◇

The proof of Proposition A.21 relies on the following lemma.

Lemma A.23. *With the same notation as above,*

(i) *If $\mathbb{E}(\xi^3) < \infty$, then there exists a constant $c_1 > 0$ such that, along any sequence of n such that $\mathbb{P}(|T| = n) > 0$,*

$$f(n) \;=\; n\,\mathbb{P}(|T| = n) \;=\; \frac{c_1}{\sqrt{n}} + O(n^{-1}).$$

(ii) *If $\mathbb{E}(\xi^2) < \infty$, then for any sequence of integers $(k_n)_{n \in \mathbb{N}}$ such that $k_n = o(\sqrt{n})$, we have*

$$\mathbb{P}\Big(|T_*^{k_n}| \geqslant n\Big) \;=\; \Theta\Big(\frac{k_n}{\sqrt{n}}\Big).$$

Proof. (i) By Minami's correspondence [22, Theorem 2], we have

$$\mathbb{P}(|T| = n) \; = \; \mathbb{P}(v(T') = n),$$

where $v(T')$ denotes the number of vertices of T', and T' is a GW(ξ') tree whose offspring distribution is

$$\xi' \; = \; \sum_{i=1}^{Z} Y_i \,,$$

with $\mathbb{P}(Z = n) = \mathbb{P}(\xi > 0)^n \, \mathbb{P}(\xi = 0)$ for all $n \geqslant 0$, and where the variables Y_i are i.i.d., independent of Z, and distributed as $(\xi - 1 \mid \xi > 0)$. It is easy to see that T' is critical and that ξ' has a finite third moment.

Next, we use the well-known fact (see e.g. [9]) that

$$n \, \mathbb{P}(v(T') = n) \; = \; \mathbb{P}(S_n = n - 1), \quad \text{with } S_n = \sum_{i=1}^{n} \xi'_i,$$

where the ξ'_i are i.i.d. copies of ξ'. Since ξ' has a finite third moment, the asymptotic expansion given in (i) follows readily from a Berry–Esseen type estimate for the local central limit theorem [23, Chapter VII, Theorem 6].

(ii) Let us fix a sequence $k_n = o(\sqrt{n})$. By construction of Kesten's tree, we have

$$\left| T_*^{k_n} \right| - 1 \; \overset{d}{=} \; \sum_{i=1}^{k_n} X_i, \tag{24}$$

where the X_i are independent and distributed as $\sum_{i=1}^{Y} |T_i|$, where Y is distributed as $\hat{\xi} - 1$ and the T_i are i.i.d. GW(ξ) trees.

Similarly to (i), using a standard local central limit theorem – this time, requiring only a finite second moment for ξ – we get $\mathbb{P}(|T| \geqslant n) \sim C_1 n^{-1/2}$ for some constant $C_1 > 0$ (for a complete proof in a more general case, see e.g. [20, Theorem 3.1]). Because $\hat{\xi} - 1$ has a finite expectation, we also have $\mathbb{P}(X_1 \geqslant n) \sim C_2 n^{-1/2}$ for the variables X_i in (24), for some constant $C_2 > 0$. Therefore, X_1 is in the domain of attraction of a (1/2)-stable distribution and we can apply Heyde's large deviation theorem [13]: for any nondecreasing sequence (x_m) such that $\sum_{i=1}^{m} X_i / x_m \to 0$ in probability, we have

$$\mathbb{P}\left(\sum_{i=1}^{m} X_i \geqslant x_n \right) \; = \; \Theta\Big(m \, \mathbb{P}(X_1 \geqslant x_n) \Big) \; = \; \Theta\Big(\frac{m}{\sqrt{x_n}} \Big).$$

Since $k_n = o(\sqrt{n})$ and since X_1 is in the domain of a (1/2)-stable distribution, we have $\sum_{i=1}^{k_n} X_i / n \to 0$ in probability; and so the equation above translates to

$$\mathbb{P}\Big(|T_*^{k_n}| \geqslant n \Big) \; = \; \mathbb{P}\left(\sum_{i=1}^{k_n} X_i \geqslant n \right) \; = \; \Theta\Big(\frac{k_n}{\sqrt{n}} \Big),$$

concluding the proof. □

Proof of Proposition A.21. Recall from (23) that for any fixed tree $\mathbf{t}$ with a distinguished leaf at height k, we have

$$\mathbb{P}\Big(T_n^k = \mathbf{t} \Big) \; = \; \mathbb{P}\Big(T_*^k = \mathbf{t} \Big) \, \frac{f(n - |\mathbf{t}| + 1)}{f(n)},$$

with $f(n) = n\mathbb{P}(|T| = n)$. This implies that, letting $(\cdot)_+$ denote the positive part,

$$\begin{aligned} d_{\mathrm{TV}}\Big(T_n^k,\, T_*^k\Big) \;&=\; \sum_{\mathbf{t}} \Big(\mathbb{P}(T_*^k = \mathbf{t}) - \mathbb{P}(T_n^k = \mathbf{t})\Big)_+ \\ &\leqslant\; \mathbb{P}\Big(|T_*^k| > \tfrac{n}{2}\Big) \;+\; \sum_{\mathbf{t}:|\mathbf{t}|\leqslant n/2} \Big(\mathbb{P}(T_*^k = \mathbf{t}) - \mathbb{P}(T_n^k = \mathbf{t})\Big)_+ \\ &=\; \mathbb{P}\Big(|T_*^k| > \tfrac{n}{2}\Big) \;+\; \mathbb{E}\left[\left(1 - \frac{f(n - |T_*^k| + 1)}{f(n)}\right)_+ \cdot\, \mathbb{1}_{\{|T_*^k|\leqslant \frac{n}{2}\}}\right]. \end{aligned} \tag{25}$$

Using Lemma A.23 (i), we see that as $n \to \infty$ and uniformly for all $\ell \leqslant n/2$ such that $f(n - \ell + 1) > 0$,

$$\frac{f(n-\ell+1)}{f(n)} \;=\; \frac{\frac{c_1}{\sqrt{n-\ell+1}} + O(n^{-1})}{\frac{c_1}{\sqrt{n}} + O(n^{-1})} \;=\; \sqrt{\frac{n}{n-\ell+1}} \;+\; O(n^{-1/2})\,.$$

Because $(\frac{n}{n-\ell+1})^{1/2} \geqslant 1$, this gives

$$\left(1 - \frac{f(n-\ell+1)}{f(n)}\right)_+ \;=\; O(n^{-1/2})\,,$$

and by plugging this into (25) we obtain

$$d_{\mathrm{TV}}\Big(T_n^k,\, T_*^k\Big) \;\leqslant\; \mathbb{P}\Big(|T_*^k| > \tfrac{n}{2}\Big) \;+\; O(n^{-1/2}).$$

Finally, since T_n has exactly n leaves, the following lower bound is immediate:

$$d_{\mathrm{TV}}\Big(T_n^k,\, T_*^k\Big) \;\geqslant\; \mathbb{P}\Big(|T_*^k| > n\Big).$$

Considering a sequence $k_n = o(\sqrt{n})$ and using Lemma A.23 (ii) readily concludes the proof of the proposition. ☐

A.6 Blowups of Galton–Watson trees: proofs

In this appendix, we prove the results stated in Section 3.3 of the main text. Most of the appendix is devoted to the proof of Theorem 3.7, but more elementary and general results about the B_2 index of blowups of (not necessarily random) trees are also given.

Throughout this appendix, we use the notation and assumptions of Appendix A.5: $T \sim \mathrm{GW}(\xi)$ is a Galton–Watson tree with offspring distribution ξ, and we assume that ξ is critical, with $\mathbb{P}(\xi = 0) > 0$ and $\mathbb{E}(\xi^3) < \infty$. We denote by T_n (resp. T_*) the corresponding tree conditioned on having n leaves (resp. Kesten tree).

We also use the notation of Section 3.3: G_n and G_* will denote blowups of T_n and T_* with respect to a fixed sequence of distributions $\nu = (\nu_k)_{k\geqslant 1}$. Finally, for any G blowup of T and for each vertex $u \in T$, we let $\Gamma_u \sim \nu_{d^+(u)}$ denote the random network associated to u in the blowup construction of G.

In the statement of the next lemma, let us write $(u_i)_{i\geqslant 0}$ for the vertices along the spine of T_*, and for each $i \geqslant 0$ let v_i be the root of Γ_{u_i}.

Lemma A.24. *Let $p_k = p_{v_k}$ denote the probability (conditional on G_*) for the directed random walk on G_*, started from the root, to reach v_k. Then,*

(i) $\mathbb{E}(p_k) = (1-\eta_0)^k$.

(ii) $p_k = O((1-\eta_0)^k)$ *almost surely.*

where $\eta_0 = \mathbb{P}(\xi = 0) > 0$.

Proof. For $k \geqslant 0$, let G_*^k denote the almost surely finite network obtained from G^* by removing all (strict) descendants of v_k – with the notation of Appendix A.5, this is a blowup of T_*^k with respect to ν. Let us now fix $k \geqslant 0$ and define D to be the out-degree of u_k, i.e. the number of leaves of Γ_{u_k}. For each $i = 1, \dots, D$, let then q_i denote the probability, conditional on Γ_{u_k}, for a random walk started from v_k to pass through the i-th leaf of Γ_{u_k}. Because Γ_{u_k} is leaf-exchangeable, we have

$$\mathbb{E}\Big(q_1 \,\Big|\, G_*^k, D\Big) \;=\; \frac{1}{D}\,,$$

and, because u_{k+1} is chosen uniformly among the children of u_k,

$$\mathbb{E}\left(\frac{p_{k+1}}{p_k} \,\middle|\, G_*^k, D\right) \;=\; \mathbb{E}\Big(q_1 \,\Big|\, G_*^k, D\Big) \;=\; \frac{1}{D}.$$

Integrating with respect to D, we get

$$\frac{1}{p_k}\,\mathbb{E}\Big(p_{k+1} \,\Big|\, G_*^k\Big) \;=\; \sum_{n \geqslant 1} n\,\mathbb{P}(\xi = n) \cdot \frac{1}{n} \;=\; \mathbb{P}(\xi \geqslant 1) \;=\; 1 - \eta_0.$$

Because the out-degrees of the vertices are i.i.d. on the spine of T_*, it follows that $(1-\eta_0)^{-k} p_k$ is a positive martingale. As a result, it converges almost surely to a nonnegative random variable. This concludes the proof. ☐

We are now ready to prove Theorem 3.7 from the main text, whose statement we recall here for convenience.

Theorem 3.7 (repeated from Section 3.3)**.** *With the notation above, assuming that ξ satisfies $\mathbb{E}(\xi) = 1$, $\mathbb{P}(\xi = 0) > 0$ and $\mathbb{E}(|\xi|^3) < +\infty$, we have:*

(i) $B_2(G_n) \to B_2(G_*)$ *in distribution.*

(ii) *For all $p \geqslant 1$, $\mathbb{E}[B_2(G_n)^p] \to \mathbb{E}[B_2(G_*)^p]$, and all these moments are finite.*

Proof. The proof of the convergence in distribution consists in building a coupling of $(G_n)_{n \geqslant 1}$ and G_* such that any subsequence of $(B_2(G_n))_{n \geqslant 0}$ has a subsequence that converges to $B_2(G_*)$ almost surely. To prove the convergence of moments, we will then show that, for all $m \geqslant 1$, the sequence $(\mathbb{E}(B_2(G_n)^m))_{n \geqslant 1}$ is bounded.

Let us describe our coupling: first, pick any sequence of integers $(k_n)_{n \geqslant 1}$ such that

$$\log\log n \;\ll\; k_n \;\ll\; n^{1/2-\varepsilon},$$

for some $\varepsilon > 0$. Then, build (G_*, T_*) as in Section 3.3: let $\mathcal{U}$ denote the Ulam–Harris tree, and let $\Gamma = (\Gamma_v^k : v \in \mathcal{U}, k \geqslant 1)$ be a family of independent phylogenetic networks such that $\Gamma_v^k \sim \nu_k$. Finally, sample T_* independently of Γ, and let G_* be the blowup obtained from (T_*, Γ).

Next, conditional on T_*, build a sequence $(G_n, T_n)_{n\geqslant 1}$ as follows: for each $n \geqslant 1$, sample a conditioned $\mathrm{GW}(\xi)$ tree T_n with n leaves, independently of Γ and in such a way that

$$\mathbb{P}\Big(T_n^{k_n} \neq T_*^{k_n}\Big) \;=\; d_{\mathrm{TV}}\Big(T_n^{k_n}, T_*^{k_n}\Big)\,,$$

where the truncated trees T_n^k and T_*^k are defined as in Appendix A.5. Finally, let G_n be the blowup obtained from T_n and Γ.

Let us now show that any subsequence of $(B_2(G_n))_{n\geqslant 1}$ has a subsequence that converges to $B_2(G_*)$ almost surely – and, therefore, in distribution. By a standard result (see e.g. [6, Theorem 2.6]), this will prove that $B_2(G_n) \to B_2(G_*)$ in distribution. First, note that since $k_n = o(\sqrt{n})$, by Proposition A.21 we have

$$d_{\mathrm{TV}}\Big(T_n^{k_n}, T_*^{k_n}\Big) \;=\; O\Big(\frac{k_n}{\sqrt{n}}\Big) \;\xrightarrow[n\to\infty]{}\; 0\,.$$

As a result, along any increasing sequence of integers there exists a subsequence, whose range we denote by A, such that

$$\sum_{n\in A} \mathbb{P}\Big(T_n^{k_n} \neq T_*^{k_n}\Big) \;<\; \infty\,.$$

Thus, by the Borel–Cantelli lemma, there exists a random n_0 such that, for all $n \in A \cap [n_0, \infty[$, we have $T_n^{k_n} = T_*^{k_n}$ – and, therefore, $G_n^{k_n} = G_*^{k_n}$.

As previously, for each k let v_k denote the distinguished leaf of G_n^k and let $p_k = p_{v_k}$ denote the probability that the directed random walk reaches it. Since on the event $\{G_n^{k_n} = G_*^{k_n}\}$, the network G_n can be obtained from G_*^k by grafting a network with at most n leaves on v_k, by Proposition 1.3 we have

$$B_2(G_*^{k_n}) \;\leqslant\; B_2(G_n) \;\leqslant\; B_2(G_*^{k_n}) \;+\; p_{k_n}\log_2 n.$$

Because $k_n \gg \log\log n$, by Lemma A.24 the term $p_{k_n}\log_2 n$ vanishes almost surely. Finally, by Proposition A.20, $B_2(G_*^{k_n}) \to B_2(G_*)$ almost surely. As a result, $B_2(G_n) \to B_2(G_*)$ along the subsequence indexed by A – concluding the proof of point (i).

It remains to show the convergence of all moments of $B_2(G_n)$. Because we have already proved the convergence in distribution, it is sufficient to show that for all $m \geqslant 1$, $\mathbb{E}(B_2(G_n)^m)$ is bounded: indeed, this implies that $(B_2(G_n)^m)_{n\geqslant 1}$ is uniformly integrable – which, together with the convergence in distribution, implies the convergence of moments (see e.g. [16, Lemma 5.11]).

Now, note that

$$B_2(G_n)^m \;\leqslant\; \mathbb{1}_{\{G_n^{k_n} \neq G_*^{k_n}\}}(\log_2 n)^m \;+\; \mathbb{1}_{\{G_n^{k_n} = G_*^{k_n}\}}\Big(B_2(G_*^{k_n}) + p_{k_n}\log_2 n\Big)^m.$$

Using $(a+b)^m \leqslant 2^{m-1}(a^m + b^m)$ for $a, b \geqslant 0$, bounding $\mathbb{1}_{\{G_n^{k_n} = G_*^{k_n}\}}$ by 1 and taking expectations, we get

$$\mathbb{E}[B_2(G_n)^m] \;\leqslant\; d_{\mathrm{TV}}(T_n^{k_n}, T_*^{k_n})(\log_2 n)^m + 2^{m-1}\Big(\mathbb{E}[B_2(G_*^{k_n})^m] + \mathbb{E}[p_{k_n}^m](\log_2 n)^m\Big).$$

Recalling that $\log\log n \ll k_n \ll n^{1/2-\varepsilon}$ and that $d_{\mathrm{TV}}(T_n^{k_n}, T_*^{k_n}) = O(k_n/\sqrt{n})$, we see that the first term on the right-hand side vanishes as $n \to \infty$. So does

$\mathbb{E}[p_{k_n}^m](\log_2 n)^m$, since $p_{k_n}^m \leqslant p_{k_n}$ and, by Lemma A.24, $\mathbb{E}(p_{k_n}) = (1-\eta_0)^{k_n}$ with $(1-\eta_0) < 1$. Finally, $B_2(G_*^{k_n}) \leqslant B_2(G_*)$. Therefore, to conclude the proof it suffices to show that $\mathbb{E}[B_2(G_*)^m]$ is finite.

Let us fix some notation: let u_k denote the k-th vertex along the spine of T_*, and let Γ_{u_k} be the corresponding network in the blowup construction of G_*. Let $\hat{\xi}_k$ be the number of leaves of Γ_{u_k}, and for $i = 1, \ldots, \hat{\xi}_k$ let $q_{k,i}$ be the probability that the directed random walk started from the root of G_* goes through the i-th leaf of Γ_{u_k}. Note that, because of the leaf-exchangeability of Γ_{u_k}, we can assume that for all $k \geqslant 0$ the spine goes through the first leaf of Γ_{u_k}, so that $q_{k,1} = p_{k+1}$. Finally, let us write $G_{k,i}$ for the finite phylogenetic network consisting of the i-th leaf of Γ_{u_k} and all of its descendants. With this notation, by the grafting property,

$$B_2(G_*) \;=\; \sum_{k\geqslant 0}\Big(p_k\, B_2(\Gamma_{u_k}) + \sum_{i=2}^{\hat{\xi}_k} q_{k,i}\, B_2(G_{k,i})\Big).$$

Therefore,

$$\mathbb{E}[B_2(G_*)^m] \;\leqslant\; 2^{m-1}\left(\mathbb{E}\left[\Big(\sum_{k\geqslant 0} p_k\, B_2(\Gamma_{u_k})\Big)^m\right] + \mathbb{E}\left[\Big(\sum_{k\geqslant 0}\sum_{i=2}^{\hat{\xi}_k} q_{k,i} B_2(G_{k,i})\Big)^m\right]\right) \tag{26}$$

To bound the first term on the right-hand side, note that since $p_k^m \leqslant p_k$ and since p_k is independent of Γ_{u_k},

$$\mathbb{E}\Big[p_k^m B_2(\Gamma_{u_k})^m\Big] \;\leqslant\; \mathbb{E}[p_k]\,\mathbb{E}\Big[B_2(\Gamma_{u_k})^m\Big]\,.$$

Using $\mathbb{E}[p_k] = (1-\eta_0)^k$ and $B_2(\Gamma_{u_k}) \leqslant \log_2 \hat{\xi}_k$, this yields

$$\mathbb{E}\Big[p_k^m B_2(\Gamma_{u_k})^m\Big] \;\leqslant\; (1-\eta_0)^k\,\mathbb{E}\Big[(\log_2 \hat{\xi})^m\Big]\,.$$

Since ξ has a finite variance, it follows that $\hat{\xi}$ has a finite mean – and therefore that $\mathbb{E}[(\log_2 \hat{\xi})^m] < \infty$. Thus, letting $\|\cdot\|_m = \mathbb{E}(|\cdot|^m)^{1/m}$ denote the L^m-norm,

$$\|p_k\, B_2(\Gamma_{u_k})\|_m \;=\; O((1-\eta_0)^{k/m})\,.$$

As a result, by the triangle inequality,

$$\Big\|\sum_{k\geqslant 0} p_k B_2(\Gamma_{u_k})\Big\|_m \;\leqslant\; \sum_{k\geqslant 0}\|p_k\, B_2(\Gamma_{u_k})\|_m \;<\; \infty\,.$$

Finally, to bound the second term in (26), note that $\sum_{k\geqslant 0}\sum_{i=2}^{\hat{\xi}_k} q_{k,i} \;=\; 1$ almost surely. Thus, Jensen's inequality gives

$$\mathbb{E}\left[\Big(\sum_{k\geqslant 0}\sum_{i=2}^{\hat{\xi}_k} q_{k,i}\, B_2(G_{k,i})\Big)^m\right] \;\leqslant\; \mathbb{E}\left[\sum_{k\geqslant 0}\sum_{i=2}^{\hat{\xi}_k} q_{k,i}\, B_2(G_{k,i})^m\right] \;=\; \mathbb{E}\Big[B_2(G)^m\Big],$$

where G is a blowup of T, the Galton–Watson tree with offspring distribution ξ (the last equality follows from the fact that the $G_{k,i}$'s are independent of the $q_{k,i}$'s and are all distributed as G). Thus, to finish the proof it suffices to note that, since $B_2(G) \leqslant \log_2|T|$, where $|T|$ denotes the number of leaves of T, we have

$$\mathbb{E}\Big[B_2(G)^m\Big] \;\leqslant\; \mathbb{E}\Big[(\log_2|T|)^m\Big].$$

Indeed, it is classic (see also Lemma A.23) that $\mathbb{P}(|T| \geqslant n) = \Theta(n^{-1/2})$, which implies that $(\log_2|T|)^m$ is integrable. This concludes the proof. □

Let us close this appendix by listing further properties of the B_2 index of blowups of trees. Although some of these properties are not related to Galton–Watson trees, the reason why we list them here is that proving them requires some of the vocabulary and notation of this Appendix, and that they all follow readily from the following elementary proposition.

Proposition A.25. *Let* $\mathbf{t}$ *be a fixed tree, and let* G *be the blowup of* $\mathbf{t}$ *with respect to some family of random networks* $\nu = (\nu_k)_{k \geqslant 1}$. *Let* X *denote the directed random walk on* $\mathbf{t}$, *and for all* $v \in \mathbf{t}$, *let* $p_v = \mathbb{P}(v \in X)$. *Then,*

$$\mathbb{E}(B_2(G)) \;=\; \sum_{v \in \mathbf{t}} p_v f(d^+(v)) \;=\; \mathbb{E}\left[\sum_{t \geqslant 0} f(d^+(X_t))\right],$$

where $d^+(v)$ *denotes the outdegree of* v *in* $\mathbf{t}$, *and* $f \colon k \mapsto \mathbb{E}[B_2(\Gamma_k)]$, *with* $\Gamma_k \sim \nu_k$ *and the convention* $f(0) = 0$.

Proof. Let ρ denote the root of $\mathbf{t}$ and, for conciseness, write $\delta := d^+(\rho)$ for its outdegree. Let $\Gamma_\rho \sim \nu_\delta$ denote the random network associated to ρ in the blowup construction of G. Finally, let $\mathbf{t}_1, \dots, \mathbf{t}_\delta$ denote the subtrees of $\mathbf{t}$ subtended by the children of ρ, and let $G_1, \dots, G_\delta$ be the corresponding blowups with respect to ν.

Since G is obtained by grafting $G_1, \dots, G_\delta$ on the leaves of Γ_ρ, by the grafting property we have

$$B_2(G) \;=\; B_2(\Gamma_\rho) \;+\; \sum_{i=1}^{\delta} q_\rho(i)\, B_2(G_i)\,, \tag{27}$$

where $q_\rho(i)$ denotes the probability that the directed random walk on Γ_ρ ends it its i-th leaf. Note that q_ρ is a random probability distribution on $\{1, \dots, \delta\}$, and that it follows from the definition of blowups that:

(i) being a deterministic function of Γ_ρ, q_ρ is independent of $(G_1, \dots, G_\delta)$;

(ii) by the leaf-exchangeability of Γ_ρ, q_ρ is exchangeable.

As a result, taking expectations in (27) we get

$$\mathbb{E}(B_2(G)) \;=\; f(\delta) \;+\; \frac{1}{\delta} \sum_{i=1}^{\delta} \mathbb{E}(B_2(G_i))\,,$$

and $\mathbb{E}(B_2(G)) = \sum_{v \in \mathbf{t}} p_v f(d^+(v))$ follows by induction. Finally, to see that this is also $\mathbb{E}[\sum_{t \geqslant 0} f(d^+(X_t))]$, it suffices to note that $p_v = \sum_{t \geqslant 0} \mathbb{P}(X_t = v)$ for any internal vertex v, and that $\sum_{v \in \mathbf{t}} \mathbb{1}_{\{X_t = v\}} = 1$ a.s. for all $t \geqslant 0$. □

By “not blowing up” the base tree $\mathbf{t}$ in Proposition A.25 (i.e. by using star trees for the networks by which internal vertices are replaced in the blowup, so that the tree is left unchanged), we immediately get the following simple expression for the B_2 index of a tree. This expression does not seem to have been pointed out previously in the literature.

Corollary A.26. *For any tree* $\mathbf{t}$, *we have*

$$B_2(\mathbf{t}) \;=\; \sum_{v \in \mathbf{t}} p_v \log_2 d^+(v)\,,$$

where $d^+(v)$ *denotes the outdegree of* v *in* $\mathbf{t}$, *and* $p_v = \Big(\prod_u d^+(u)\Big)^{-1}$, *where the product runs over the vertices on the path from the root of* $\mathbf{t}$ *to* v, *excluding* v.

Because $f(k) \leqslant \log_2 k$, Proposition A.25 also has the following corollary.

Corollary A.27. *Let* $\mathbf{t}$ *be a fixed tree, and let* G *be a blowup of* $\mathbf{t}$*. Then,*

$$\mathbb{E}(B_2(G)) \;\leqslant\; B_2(\mathbf{t})\,.$$

Note however that the inequality in Corollary A.27 only holds after integrating with respect to the blowup procedure. In particular, if T is a random tree and G is a blowup of T, then $B_2(T)$ does not necessarily stochastically dominate $B_2(G)$. To see this, let $T = \mathbf{t}$ be deterministic, and note that it is then possible to have $B_2(G) > B_2(\mathbf{t})$. This implies that, in that case, there is not monotone coupling of $B_2(G)$ and $B_2(T)$.

Nevertheless, as the next corollary shows, $B_2(T)$ is second-order stochastically dominant over $B_2(G)$. Very loosely speaking, this means that – in addition to being no smaller in expectation – $B_2(T)$ is more predictable than $B_2(G)$.

Corollary A.28. *Let* T *be a tree, and let* G *be a blowup of* T*. Then,* $B_2(T)$ *is second-order stochastically dominant over* $B_2(G)$*, that is: for every nondecreasing concave function* φ*,*

$$\mathbb{E}[\varphi(B_2(G))] \;\leqslant\; \mathbb{E}[\varphi(B_2(T))]\,.$$

Proof. Let φ be a nondecreasing concave function. First, by Corollary A.27, we have $\mathbb{E}(B_2(G)\,|\,T) \leqslant B_2(T)$, and therefore

$$\varphi\Big(\mathbb{E}(B_2(G)\,|\,T)\Big) \;\leqslant\; \varphi(B_2(T))\,.$$

As a result, by Jensen's inequality,

$$\mathbb{E}\Big[\varphi(B_2(G))\Big] \;=\; \mathbb{E}\Big[\mathbb{E}[\varphi(B_2(G))\,|\,T]\Big] \;\leqslant\; \mathbb{E}\Big[\varphi\Big(\mathbb{E}[B_2(G)\,|\,T]\Big)\Big] \;\leqslant\; \mathbb{E}\Big[\varphi(B_2(T))\Big]\,,$$

concluding the proof. ☐

Finally, note that Proposition A.25 yields a simple derivation of the expression that was already given in Theorem 3.8 for the expected value of the B_2 index of blowups of Galton–Watson trees. Before detailing this, let us briefly compare the two approaches: recall that the proof of Theorem 3.8 used the recursive structure of Galton–Watson trees (i.e. the branching property) together with the independence between the base tree and the networks used in the blowup procedure to get a distributional equation for $B_2(G)$. In a way, Proposition A.25 isolates the part of the proof of Theorem 3.8 where the independence between the base tree and the networks is used, and in the proof below we use the branching property to recover the expression of Theorem 3.8; what justifies presenting the two proofs is that the way the branching property is used is different.

Corollary A.29. *Let* T *be a Galton–Watson tree with offspring distribution* ξ *and let* G *be a blowup of* T *with respect to a family of networks* $\nu = (\nu_k)_{k \geqslant 1}$*. Then,*

$$\mathbb{E}(B_2(T)) \;=\; \frac{\mathbb{E}(f(\xi))}{\mathbb{P}(\xi = 0)}\,,$$

where $f(0) = 0$ *and, for* $k \geqslant 1$*,* $f(k) = \mathbb{E}[B_2(\Gamma_k)]$*, where* $\Gamma_k \sim \nu_k$*, and with the convention* $0/0 = 0$*.*

Proof. First, let us deal with degenerate cases: when $\mathbb{P}(\xi = 1) = 1$, the tree T is an infinite path, and therefore G has exactly one end – which implies $B_2(G) = 0$. Since $f(1) = 0$, the proposition holds with the convention $0/0 = 0$. In the case where ξ is not almost surely equal to 1 and where $\mathbb{P}(\xi = 0) = 0$, we have $B_2(G) = +\infty$ and the proposition also holds.

Now, assume that $\mathbb{P}(\xi = 0) > 0$. Note that, by the branching property,

$$\sum_{t \geqslant 0} f(d^+(X_t)) \;\overset{d}{=}\; \sum_{i=1}^{N} f(\xi_i^+)$$

$(\xi_i^+)_{i \geqslant 1}$ are i.i.d. copies of $(\xi \mid \xi > 0)$, the variable ξ conditioned to be positive, and where N is a geometric variable on $\{0, 1, \ldots\}$ with parameter $\eta_0 := \mathbb{P}(\xi = 0)$ that is independent of $(\xi_i^+)_{i \geqslant 1}$. As a result, by Wald's formula,

$$\mathbb{E}(B_2(T)) \;=\; \mathbb{E}(N)\, \mathbb{E}[f(\xi^+)] \;=\; \frac{1 - \eta_0}{\eta_0} \cdot \frac{\mathbb{E}[f(\xi) \mathbb{1}_{\{\xi > 0\}}]}{1 - \eta_0} \;=\; \frac{\mathbb{E}[f(\xi)]}{\eta_0}\,,$$

because $f(0) = 0$. This concludes the proof. □