Check for updates

J. Appl. Probab. **62,** 1089–1104 (2025)
*doi:*10.1017/jpr.2025.14

ERGODIC THEOREM FOR BRANCHING MARKOV CHAINS INDEXED BY TREES WITH ARBITRARY SHAPE

JULIEN WEIBEL ,* *Institut Denis Poisson and CERMICS*

Abstract

We prove an ergodic theorem for Markov chains indexed by the Ulam–Harris–Neveu tree over large subsets with arbitrary shape under two assumptions: (i) with high probability, two vertices in the large subset are far from each other, and (ii) with high probability, those two vertices have their common ancestor close to the root. The assumption on the common ancestor can be replaced by some regularity assumption on the Markov transition kernel. We verify that these assumptions are satisfied for some usual trees. Finally, with Markov chain Monte Carlo considerations in mind, we prove that when the underlying Markov chain is stationary and reversible, the Markov chain, that is the line graph, yields minimal variance for the empirical average estimator among trees with a given number of nodes. In doing so, we prove that the Hosoya–Wiener polynomial is minimized over $[-1, 1]$ by the line graph among trees of a given size.

Keywords: Tree-indexed Markov chain; ergodic theorem; Bienaymé–Galton–Watson trees; minimal variance

2010 Mathematics Subject Classification: Primary 60J05; 60J80
Secondary 60F25

1. Introduction

Branching Markov processes, which are a generalization of Markov chains where processes are indexed by trees, are useful for describing the evolution and growth of a population. Limit theorems, such as the law of large numbers (sometimes also called the ergodic theorem in Markovian contexts), are important tools for studying properties of a population such as the distribution of traits. The law of large numbers for branching Markov processes has been studied with both a discrete or continuous state space [4, 5] for set indicator functions, and thus for continuous bounded functions by the portmanteau theorem. To study cellular aging, a more general version of the strong law of large numbers for a wider class of test functions and for non-independent daughter cells was proved in [10]. This wider class of admissible test functions depends on the transition kernel of the branching Markov process considered; e.g., [10] focuses on one example with an autoregressive Gaussian transition kernel and with continuous and polynomially growing functions. See also [9] for an extension of [10] to bounded-degree Bienaymé–Galton–Watson trees and [7] for an extension to time-varying environments and trait-dependent offspring distribution. In this article, we present an ergodic theorem for a wide

Received 8 April 2024; accepted 20 January 2025.
* Postal address: Institut Denis Poisson, Université d'Orléans, Université de Tours, CNRS, France and CERMICS, École des Ponts, France. Email: julien.weibel@normalesup.org

© The Author(s), 2025. Published by Cambridge University Press on behalf of Applied Probability Trust.

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press

class of test functions as in [10], and for branching Markov processes where reproduction is independent from individual traits but where the genealogical tree of the population can have an arbitrary shape.

A branching Markov process $X = (X_u, u \in T)$ with values in a metric space $\mathcal{X}$ is a random process indexed by a rooted tree T with the Markov property: sibling nodes take independent and identically distributed values that depend only on the value of their parent node. Without loss of generality we may choose T to be the rooted Ulam–Harris–Neveu tree $\mathbb{T}^\infty = \cup_{n\in\mathbb{N}}(\mathbb{N}^*)^n$. See Definition 1 for a complete formal definition.

For simplicity, in this introduction we restrict ourselves to the case where the transition kernel Q of the branching Markov process X is *ergodic* (resp. *uniformly ergodic*); that is, Q has a unique invariant measure μ and, for any continuous bounded function f on $\mathcal{X}$, we have for all $x \in \mathcal{X}$ that $\lim_{n\to\infty} |Q^n f(x) - \langle \mu, f\rangle| = 0$ (resp. $\lim_{n\to\infty} \sup_{x\in\mathcal{X}} |Q^n f(x) - \langle \mu, f\rangle| = 0$).

For a finite (non-empty) subset $A \subset \mathbb{T}^\infty$ and some function f on $\mathcal{X}$, we define the normalized empirical average as $\bar{M}_A(f) = |A|^{-1} \sum_{u\in A} f(X_u)$. Our goal is to study the asymptotic behavior of the normalized empirical average when the averages are performed on a sequence $(A_n)_{n\in\mathbb{N}}$ of finite subsets of $\mathbb{T}^\infty$ whose size goes to infinity. For instance, the averaging set A_n can be the nth generation G_n of a tree T, or T_n the tree T up to generation n. In the rest of the article we will always use this definition for G_n and T_n.

To this end, we need a geometrical assumption on the sequence of finite subsets $(A_n)_{n\in\mathbb{N}}$, which states that vertices are far away from each other with high probability.

Assumption 1. (Geometrical.) *Let $A_n \subset \mathbb{T}^\infty$ for $n \in \mathbb{N}$ be finite (non-empty) subsets, and let U_n and V_n be independent and uniformly sampled elements of A_n. Denoting by d the graph distance on $\mathbb{T}^\infty$, for all $k \in \mathbb{N}$ we have*

$$\mathbb{P}(d(U_n, V_n) \le k) = |A_n|^{-2} \sum_{u,v\in A_n} \mathbf{1}_{\{d(u,v)\le k\}} \underset{n\to\infty}{\longrightarrow} 0.$$

We stress that Assumption 1 implies that $\lim_{n\to\infty} |A_n| = \infty$.

We either need to assume that Q is uniformly ergodic, or that Q is ergodic and the sequence $(A_n)_{n\in\mathbb{N}}$ satisfies the following condition stating that the last common ancestor of two vertices is near the root with high probability. Denote by $h(u)$ the height of a vertex u and by $u \wedge v$ the common ancestor of two vertices u and v (see Section 2.1).

Assumption 2. (Ancestral) *For all $n \in \mathbb{N}$, let U_n and V_n be independent and uniformly sampled elements of A_n. The sequence of random variables $(h(U_n \wedge V_n))_{n\in\mathbb{N}}$ is tight, that is, for every $\varepsilon > 0$ there exists $k \in \mathbb{N}$ such that $\mathbb{P}(h(U_n \wedge V_n) > k) < \varepsilon$ for n large enough.*

Note that Assumptions 1 and 2 are similar to [7, Assumptions 2.(b) and 2.(a)] in the case where A_n is the nth generation of the tree.

Remark 1. (*Some sufficient conditions for Assumptions 1 and 2; see Section 3*) Assumption 1 is always satisfied for Cayley and Bethe trees, and for bounded-degree trees (see Lemma 4) as long as $\lim_{n\to\infty} |A_n| = \infty$. Assumptions 1 and 2 are satisfied for spherically symmetric trees when $A_n = G_n$ (see Lemma 6). In Lemma 7 , we prove that Assumptions 1 and 2 are satisfied for super-critical Bienaymé–Galton–Watson trees conditioned on non-extinction when $A_n = G_n$ or T_n.

We can now formulate the ergodic theorem for branching Markov processes on trees with arbitrary shape. In Section 2, we prove Theorem 2, a more general version of this theorem.

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press

Theorem 1. (Ergodic theorem for Markov processes on trees with arbitrary shape.) *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of finite subsets of $\mathbb{T}^\infty$ that satisfies Assumption 1. Let X be a branching Markov process indexed by $\mathbb{T}^\infty$ with values in $\mathcal{X}$ whose transition kernel Q is ergodic. Assume that either Q is uniformly ergodic or that $(A_n)_{n\in\mathbb{N}}$ satisfies Assumption 2. Then, for every continuous bounded function f on $\mathcal{X}$,*

$$\bar{M}_{A_n}(f) = |A_n|^{-1} \sum_{u\in A_n} f(X_u) \xrightarrow[n\to\infty]{L^2} \langle \mu, f\rangle.$$

Remark 2. We discuss the main difference between Theorem 1 and the law of large numbers for a branching Markov process found in [10]. The results in [10] apply to Markov processes where daughter nodes can have non-independent distributions when conditioning on their mother, whereas in our case they must be independent. In exchange, our results allow for more flexibility on the shape of the population's genealogical tree: for instance, more flexibility on the number of children of each node (including Bienaymé–Galton–Watson trees with unbounded degree), or even allowing the number of children of a node to grow over time (e.g. the degree of the root can increase as $\log n$, i.e. slow condensation). Our results could also be applied to random trees with population-wise interactions (e.g. competition where reproduction rate decreases when population size increases). Moreover, in our results, the empirical average can be performed on a wide variety of (possibly random) subsets of the tree, and not only to the nth generation. As an example, our results could be applied to a random subset of size $\log n$ of the nth generation G_n of a super-critical Bienaymé–Galton–Watson tree chosen uniformly at random. Also note that in our results, assumptions can be verified separately for the transition kernel Q and the population's genealogical tree.

Lastly, motivated by Markov chain Monte Carlo considerations, in Section 4 we study the variance of the empirical average estimator $\bar{M}_A(f)$ and its dependence on the shape of A. We perform exact variance computation in the case where the transition kernel Q induces a self-adjoint compact operator on $L^2(\mu)$, which proves the following proposition about non-asymptotic variance comparison.

Proposition 1. (The line graph has minimal variance.) *Let μ be an invariant measure for Q, and assume that the transition kernel Q induces a self-adjoint compact operator on $L^2(\mu)$. Let X be a branching Markov process on $\mathbb{T}^\infty$ with transition kernel Q and initial distribution ν. Let f be a non-constant function in $L^2(\mu)$.*

When $\nu = \mu$, $\mathbb{E}[\bar{M}_A(f)] = \langle \mu, f\rangle$ for any finite subset $A \subset \mathbb{T}^\infty$, and thus the empirical average estimator has no bias. Moreover, the minimum of the map $A \mapsto \mathrm{Var}(\bar{M}_A(f))$ among subtrees of $\mathbb{T}^\infty$ with a given cardinal n is achieved by the line graph tree (i.e. the Markov chain).

Furthermore, when $n \geq 5$, the line graph is the only subtree of size n achieving this minimum if and only if $f \notin \mathrm{Ker}(Q) \oplus \mathrm{Ker}(Q - I) \oplus \mathrm{Ker}(Q + I)$.

Note that as Q is a Markov kernel, its spectrum as an $L^2(\mu)$-operator is a subset of $[-1, 1]$. Note that when $f \in \mathrm{Ker}(Q) \oplus \mathrm{Ker}(Q - I)$, then the value of $\mathrm{Var}(\bar{M}_A(f))$ does not depend on the shape of the tree A. Also note that when $f \in \mathrm{Ker}(Q + I)$, then the value of $\mathrm{Var}(\bar{M}_A(f))$ is minimal among subtrees of size n when A has a balanced bipartite 2-coloring, and for $n \geq 5$, the line graph is not the only tree with a balance bipartite 2-coloring.

Hence, if we want to approximate $\langle \mu, f\rangle$, using a branching Markov chain does not improve the rate of convergence compared to a standard Markov chain.

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press

The proof of Proposition 1 relies on decomposing the function f on a basis of eigenvectors of Q; the problem then reduces to minimization among trees of a given size of the Hosoya–Wiener polynomial $H_A(\alpha) = \sum_{u,v\in A} \alpha^{d(u,v)}$ for some $\alpha \in [-1, 1]$. This leads us to prove the following lemma stating that the line graph tree achieves this minimum.

Lemma 1. (The line graph minimizes the Hosoya–Wiener polynomial.) *Let $\alpha \in [-1, 1]$ and $n \in \mathbb{N}^*$. Then the minimum of the map $A \mapsto H_A(\alpha)$ among trees with given cardinal n is achieved by the line graph tree. Furthermore, when $\alpha \in (-1, 0) \cup (0, 1)$, the line graph tree of size n is the only tree achieving this minimum.*

Note that this result was already proved for $\alpha \in [0, 1]$ in [8, Theorem 9], where the proof relies heavily on the monotonicity of the function $d \mapsto \alpha^d$ (and more general results were proved in [12, Theorem 2] and [13, Corollary 2.8] for Wiener-type indices of the form $W_f(A) = \sum_{u,v\in A} f(d(u, v))$ where f is a monotonic function). Hence, the novelty of Lemma 1 is for the case of $\alpha \in [-1, 0)$ where the function $d \mapsto \alpha^d$ is non-monotonic, and our proof relies on considering several cases depending on the tree structure.

2. Main theorem

2.1. Notation

Let $\mathbb{T}^\infty = \cup_{n\in\mathbb{N}}(\mathbb{N}^*)^n$ denote the Ulam–Harris–Neveu tree, and denote by ∂ its root, i.e. the empty word.

Let $u \in \mathbb{T}^\infty$ be a vertex. If u is distinct from the root, we denote by $\mathrm{p}(u)$ its parent vertex. We denote by $h(u)$ its height, i.e. the number of edges separating u from the root ∂. (The height of the root ∂ is zero.) For two vertices $u, v \in \mathbb{T}^\infty$, we denote by $u \wedge v$ the latest common ancestor of u and v, and by $d(u,v)$ the graph-distance between u and v in $\mathbb{T}^\infty$, that is, $d(u, v) = h(u) + h(v) - 2h(u \wedge v)$. Let $X = (X_u, u \in \mathbb{T}^\infty)$ be a stochastic process with values in a metric space $\mathcal{X}$.

Definition 1. (*Markov process.*) The stochastic process X is called a *(branching) Markov process* with transition kernel Q and initial distribution ν if, for any finite subtree $T \subset \mathbb{T}^\infty$ with $\partial \in T$, $\mathbb{P}\big(\bigcap_{u\in T}\{X_u \in \mathrm{d}x_u\}\big) = \nu(\mathrm{d}x_\partial) \prod_{u\in T\setminus\{\partial\}} Q(x_{\mathrm{p}(u)};\mathrm{d}x_u)$.

We denote by νQ^n the distribution of a vertex in the nth generation. For a (Borel) function f, define the function $Qf : x \in \mathcal{X} \mapsto \int f(y)\, Q(x; \mathrm{d}y)$ when the expression makes sense. For a measure μ and a Borel function f on $\mathcal{X}$, we write $\mu f = \langle \mu, f\rangle = \int_{\mathcal{X}} f\, \mathrm{d}\mu$. For the rest of this section, we fix ν and Q.

2.2. Statement of the main result

Firstly, we need some assumptions on the Borel function f with which we perform the empirical averages.

Assumption 3. (Boundedness and convergence.) *Let f be a Borel function on $\mathcal{X}$ such that:*

(i) $\sup_{n\in\mathbb{N}} \nu Q^n(f^2) < \infty$;

(ii) *there exists a constant $c_f \in \mathbb{R}$ such that $\lim_{n\to\infty} \nu Q^k((Q^n f - c_f)^2) = 0$ for all $k \in \mathbb{N}$.*

Note that if f satisfies Assumption 3, then so does $f - c$ for any $c \in \mathbb{R}$, thus we may assume that $c_f = 0$ when necessary. Also note that, using the Cauchy–Schwarz and Jensen's

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press

inequalities, Assumption 3(i) implies that $Q^n f$, $Q^n f^2$, and $Q^k(Q^n f \times Q^m f)$ (with $n, m, k \in \mathbb{N}$) are well-defined and finite ν-almost everywhere, and are ν-integrable.

Note that when Q is ergodic, then Assumption 3 is satisfied by any continuous bounded function f on $\mathcal{X}$, and we get $c_f = \langle \mu, f \rangle$, where μ is the unique invariant measure of Q. Also note that if F is a subspace of Borel functions on $\mathcal{X}$ that satisfy [10, Assumptions (i)–(vi), pp. 11–12], then any function $f \in F$ satisfies Assumption 3 with $c_f = \langle \mu, f \rangle$.

For a finite subset $A \subset \mathbb{T}^\infty$ and a Borel function f, we define the empirical sum $M_A(f) = \sum_{u \in A} f(X_u)$ and the empirical average $\bar{M}_A(f) = |A|^{-1} \sum_{u \in A} f(X_u)$, where $|A|$ is the cardinal of the set A. Let $(A_n)_{n \in \mathbb{N}}$ be a sequence of finite subsets of $\mathbb{T}^\infty$ on which we perform the empirical averages in the ergodic theorem on trees with arbitrary shape. Recall the geometrical Assumptions 1 and 2.

Contrary to the case of the binary tree considered in [10], Assumption 2 is not always satisfied for a sequence $(A_n)_{n \in \mathbb{N}}$ of finite subsets of $\mathbb{T}^\infty$ with arbitrary shape (e.g. in the case of the line graph, i.e. the Markov chain). Thus, to prove the ergodic theorem for branching Markov chains, as an alternative to the ancestral Assumption 2, we also consider the following conditions on the ergodicity of the transition kernel Q.

Assumption 4. (Stronger ergodicity.) *Assume that any of the following conditions holds:*

(*i*) $\nu = \mu$ *is an invariant measure of* Q.

(*ii*) *There is convergence in total variation* $\lim_{n \to \infty} \|\nu Q^n - \mu\|_{\mathrm{TV}} = 0$ *to some invariant measure* μ *(for* Q*), the function* f *is bounded, and we have* $\lim_{n \to \infty} \mu(Q^n f - c_f)^2 = 0$*, where* c_f *is the same constant as in Assumption 3(ii).*

(*iii*) *The transition kernel* Q *satisfies a uniformly ergodic assumption (with* μ *as its unique invariant measure): there exists a non-negative Borel function* g *on* $\mathcal{X}$ *with* $\sup_{k \in \mathbb{N}} \nu Q^k g^2 < \infty$ *and a sequence of positive numbers* $(a_n)_{n \in \mathbb{N}}$ *that converges to zero such that, for all* $n \in \mathbb{N}$, $|Q^n f - \langle \mu, f \rangle| \leq a_n\, g$.

Note that, using dominated convergence with domination by g, Assumption 4(iii) implies Assumption 3(ii) with $c_f = \langle \mu, f \rangle$. Also note that when Assumption 3 holds and either Assumption 4(i) or 4(ii) holds, then we have $c_f = \langle \mu, f \rangle$ in Assumption 3 (indeed, using Jensen's inequality, $(\langle \mu, f \rangle - c_f)^2 = \limsup_{n \to \infty} (\mu Q^n f - c_f)^2 \leq \limsup_{n \to \infty} \mu(Q^n f - c_f)^2 = 0$).

We can now formulate the ergodic theorem for branching Markov processes on trees with arbitrary shape.

Theorem 2. (Ergodic theorem for Markov processes on trees with arbitrary shape.) *Let* $(A_n)_{n \in \mathbb{N}}$ *be a sequence of finite subsets of* $\mathbb{T}^\infty$ *that satisfies Assumption 1. Let* X *be a branching Markov process on* $\mathbb{T}^\infty$ *with transition kernel* Q *and initial distribution* ν*. Let* f *be a Borel function on* $\mathcal{X}$ *that satisfies Assumption 3. Furthermore, assume that either Assumption 2 or 4 holds. Then*

$$\bar{M}_{A_n}(f) = |A_n|^{-1} \sum_{u \in A_n} f(X_u) \xrightarrow[n \to \infty]{L^2(\nu)} c_f.$$

In particular, when the transition kernel Q is ergodic and f is a continuous bounded function, then Assumption 3 is satisfied and $c_f = \langle \mu, f \rangle$, where μ is the unique invariant measure of Q. If, furthermore, Q is uniformly ergodic, then Assumption 4(iii) holds. Hence, Theorem 2 implies Theorem 1.

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press

Proof. Up to replacing f by $f - c_f$, assume that $c_f = 0$. For all $n \in \mathbb{N}$, we have

$$\mathbb{E}[\bar{M}_{A_n}(f)^2] = |A_n|^{-2} \sum_{u,v \in A_n} \mathbb{E}[f(X_u)f(X_v)]. \tag{1}$$

Note that, for $u, v \in \mathbb{T}^\infty$,

$$\mathbb{E}[f(X_u)f(X_v)] = \nu Q^{h(u\wedge v)}(Q^{d(u\wedge v,u)}f \times Q^{d(u\wedge v,v)}f). \tag{2}$$

Set $C = \sup_{n\in\mathbb{N}} \nu Q^n f^2 < \infty$, which is finite by Assumption 3(i). Hence, using the Cauchy–Schwarz and Jensen's inequalities, for all $k, \ell, m \in \mathbb{N}$ we have

$$\begin{aligned} |\nu Q^k(Q^\ell f \times Q^m f)| &\le (\nu Q^k (Q^{\min(\ell,m)}f)^2 \times \nu Q^k (Q^{\max(\ell,m)}f)^2)^{1/2} \\ &\le (\nu Q^{k+\min(\ell,m)} f^2 \times \nu Q^k (Q^{\max(\ell,m)}f)^2)^{1/2} \\ &\le \sqrt{C} \times \sqrt{\nu Q^k (Q^{\max(\ell,m)}f)^2}. \end{aligned} \tag{3}$$

Define the distance $\tilde{d}$ on $\mathbb{T}^\infty$ as $\tilde{d}(u, v) = \max(d(u, u\wedge v), d(v, u\wedge v)) = \max(h(u), h(v)) - h(u\wedge v)$. Note that we have $d/2 \le \tilde{d} \le d$, thus Assumption 1 is equivalent to: for all $k \in \mathbb{N}$, $\lim_{n\to\infty} \mathbb{P}(\tilde{d}(U_n, V_n) \le k) = 0$. Then, as a consequence of (1), (2), and (3), we get

$$\begin{aligned} \mathbb{E}[\bar{M}_{A_n}(f)^2] &\le \sqrt{C} \times |A_n|^{-2} \sum_{u,v\in A_n} (\nu Q^{h(u\wedge v)} (Q^{\tilde{d}(u,v)}f)^2)^{1/2} \\ &\le \sqrt{C} \times \left(|A_n|^{-2} \sum_{u,v\in A_n} \nu Q^{h(u\wedge v)} (Q^{\tilde{d}(u,v)}f)^2 \right)^{1/2}, \end{aligned}$$

where we used Jensen's inequality in the last inequality. Hence, to conclude the proof it is enough to prove that the following holds:

$$\mathbb{E}[\nu Q^{h(U_n\wedge V_n)} (Q^{\tilde{d}(U_n,V_n)}f)^2] = |A_n|^{-2} \sum_{u,v\in A_n} \nu Q^{h(u\wedge v)} (Q^{\tilde{d}(u,v)}f)^2 \underset{n\to\infty}{\longrightarrow} 0. \tag{4}$$

As Assumption 2 (resp. Assumption 4) holds, using Lemma 2 (resp. Lemma 3) we get that (4) holds, which concludes the proof. □

Recall that, from Assumption 3, we know that $\lim_{n\to\infty} \nu Q^k (Q^n f)^2 = 0$ for all $k \in \mathbb{N}$, and that (4) adds some uniformity in k. In the next lemma, we check that the ancestral Assumption 2, which allows for small values $h(U_n \wedge V_n)$ with high probability, implies (4).

Lemma 2. *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of finite subsets of $\mathbb{T}^\infty$ that satisfies Assumption 1. Let f be a function on $\mathcal{X}$ that satisfies Assumption 3. Then, Assumption 2 implies (4).*

Proof. Without loss of generality, assume that $c_f = 0$. Set $C = \sup_{j\in\mathbb{N}} \nu Q^j f^2 < \infty$, which is finite by Assumption 3(i). Thus, for $k, m \in \mathbb{N}$, using Jensen's inequality,

$$\nu Q^k (Q^m f)^2 \le \nu Q^{k+m} f^2 \le C < \infty. \tag{5}$$

Let $\varepsilon > 0$. Using Assumption 2, there exists $K \in \mathbb{N}$ such that, for $n \in \mathbb{N}$ large enough, we have $\mathbb{P}(h(U_n \wedge V_n) > K) < \varepsilon$. Using Assumption 3(ii), let $M \in \mathbb{N}$ be such that, for all $m \ge M$ and

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press

for all $k \le K$, $\nu Q^k (Q^m f)^2 < \varepsilon$. Using Assumption 1, for n large enough we have $\mathbb{P}(\tilde{d}(U_n, V_n) < M) < \varepsilon$. Hence, using (5), for n large enough we get

$$\mathbb{E}[\nu Q^{h(U_n \wedge V_n)} (Q^{\tilde{d}(U_n, V_n)} f)^2] \le 2C\varepsilon + \max_{k \le K} \sup_{m \ge M} \nu Q^k (Q^m f)^2 < (1 + 2C)\varepsilon.$$

This being true for all $\varepsilon > 0$, (4) holds, which concludes the proof. □

The following lemma states that the stronger ergodic Assumption 4 implies (4); its proof is similar to the proof of Lemma 2 and is left to the reader.

Lemma 3. *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of finite subsets of $\mathbb{T}^\infty$ that satisfies Assumption 1. Let f be a function on $\mathcal{X}$ that satisfies Assumption 3. Then, Assumption 4 implies (4).*

3. Examples satisfying Assumptions 1 and 2

We now give some common examples of trees for which Assumptions 1 and 2 are satisfied.

We denote by T an arbitrary infinite tree rooted at some vertex ∂. In this section, all the trees we consider are locally finite (i.e. all nodes have finite degree). For $n \in \mathbb{N}$, we denote by G_n the nth generation of T, that is, the set of vertices at distance n from the root, and we denote by $T_n = \cup_{k=0}^n G_k$ the tree up to generation n. For a vertex $u \in T$, we denote by $T(u)$ the subtree of T rooted at u and composed of all descendants of u.

3.1. Some simple deterministic trees

Firstly, Assumption 1 is always satisfied on trees with bounded vertex degrees, in particular for Cayley and Bethe trees (i.e. trees where each non-leaf vertex has out-degree D for some $D \ge 1$, except for the root of a Bethe tree which has out-degree $D + 1$).

Lemma 4 (Bounded-degree trees.) *Let $D \ge 2$, and let T be the infinite rooted complete D-ary tree (i.e. the tree where each vertex has D children). Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of finite (non-empty) subsets of T such that $\lim_{n\to\infty} |A_n| = \infty$. Then the sequence $(A_n)_{n\in\mathbb{N}}$ satisfies Assumption 1.*

Proof. Let $k \in \mathbb{N}$. For every vertex $u \in T$, the ball $B_T(u, k)$ of radius k and center u has cardinal upper bounded by $c_k = \sum_{j=0}^k (D + 1)^j$. Hence,

$$\frac{1}{|A_n|^2} \sum_{u,v \in A_n} \mathbf{1}_{\{d(u,v) \le k\}} = \frac{1}{|A_n|^2} \sum_{u \in A_n} |B_T(u, k)| \le \frac{1}{|A_n|} c_k.$$

This implies that Assumption 1 is satisfied. □

The following counterexample shows that Assumption 2 is not always satisfied on a bounded degree tree.

Example 1. Let T be the infinite tree where each vertex has out-degree $D \ge 2$. Set $A_n = T(u_n) \cap G_{2n}$, the nth descendants of u_n, where $u_n = 1 \cdots 1$ (n times) is the left-most nth descendant of the root. Then $|A_n| = D^n \to_{n\to\infty} \infty$ and, by Lemma 4 , the sequence $(A_n)_{n\in\mathbb{N}}$ satisfies Assumption 1. However, $\mathbb{P}(h(U_n \wedge V_n) \ge n) = 1$ for all $n \in \mathbb{N}$, which implies that the sequence $(A_n)_{n\in\mathbb{N}}$ does not satisfy Assumption 2.

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press

When the vertices in A_n are far from the root with high probability (e.g. when $A_n = G_n$) and have their common ancestor close to the root, then Assumption 1 is satisfied.

Lemma 5. (A_n far from the root.) *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of finite subsets of $\mathbb{T}^\infty$. For every $n \in \mathbb{N}$, let U_n and V_n be independent and uniformly sampled elements of A_n. Assume that, for every $k \in \mathbb{N}$, $\lim_{n\to\infty} \mathbb{P}(h(U_n) \le k) = 0$ and that Assumption 2 holds. Then the sequence $(A_n)_{n\in\mathbb{N}}$ satisfies Assumption 1.*

Proof. Recall that $d(U_n, V_n) = h(U_n) + h(V_n) - 2h(U_n \wedge V_n)$. Thus, for $k \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}(d(U_n, V_n) \le 2k) &\le 2\,\mathbb{P}(h(U_n) - h(U_n \wedge V_n) \le k)\\ &\le 2\,\mathbb{P}(h(U_n \wedge V_n) > k) + 2\,\mathbb{P}(h(U_n) \le 2k),\end{aligned}$$

where both terms in the upper bound go to zero as $n \to \infty$. Hence, the sequence $(A_n)_{n\in\mathbb{N}}$ satisfies Assumption 1. □

We say a tree T is spherically symmetric (sometimes also called a generalized Bethe tree) if, for all $n \in \mathbb{N}$, every vertex of height n in T has the the same out-degree D_n. When we choose $A_n = G_n$ the nth generation for all $n \in \mathbb{N}$, Assumptions 1 and 2 are always satisfied on spherically symmetric trees.

Lemma 6. (Spherically symmetric trees.) *Let T be an infinite spherically symmetric tree such that $\lim_{n\to\infty} |G_n| = \infty$. Then the sequence $(G_n)_{n\in\mathbb{N}}$ satisfies Assumptions 1 and 2.*

Proof. Thanks to Lemma 5 , we only need to prove that Assumption 2 is true. For all $n \in \mathbb{N}$, denote by D_n the out-degree for all vertices of height n. As $\lim_{n\to\infty} |G_n| = \infty$, we have $D_n > 1$ for infinitely many values of n. Let U_n and V_n be independent random vertices uniformly distributed over G_n. Using the Ulam–Harris–Neveu tree notation, write $U_n = U_{(1)} \cdots U_{(n)}$ and $V_n = V_{(1)} \cdots V_{(n)}$, where the random variables $U_{(1)}, \ldots, U_{(n)}, V_{(1)}, \ldots, V_{(n)}$ are independent, with $U_{(i)}$ and $V_{(i)}$ uniformly distributed over the set $\{1, \ldots, D_{i-1}\}$. Thus, for all $k \in \mathbb{N}$ and $n \ge k$, as $d(U_n, V_n) = 2n - 2h(U_n \wedge V_n)$, we have

$$\mathbb{P}(h(U_n \wedge V_n) \ge k) = \mathbb{P}(U_{(i)} = V_{(i)} \text{for all } i \in \{1, \ldots, k\}) = \prod_{i=0}^{k-1} \frac{1}{D_i},$$

where the right-hand side goes to 0 as $k \to \infty$. This implies that Assumption 2 is satisfied, and thus concludes the proof. □

3.2. Super-critical Bienaymé–Galton–Watson trees

To apply the ergodic Theorem 2 to a random sequence $(A_n)_{n\in\mathbb{N}}$ of subsets of $\mathbb{T}^\infty$ (independent of the Markov process indexed by $\mathbb{T}^\infty$), we need to verify that, almost surely (a.s.), the sequence $(A_n)_{n\in\mathbb{N}}$ satisfies Assumption 1 (and possibly Assumption 2). Note that in this case, Assumptions 1 and 2 should be considered conditionally on A_n, in particular the random vertices U_n and V_n are independent and uniformly distributed over A_n conditionally on A_n.

In this subsection we consider the case where T is a super-critical Bienaymé–Galton–Watson tree whose offspring distribution $\mathcal{P}$ on $\mathbb{N}$ has finite mean $m > 1$ and finite second moment, and is conditioned on non-extinction. The following lemma states that, a.s., both the sequences $(G_n)_{n\in\mathbb{N}}$ and $(T_n)_{n\in\mathbb{N}}$ satisfy Assumptions 1 and 2.

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press

Lemma 7. (Super-critical Bienaym–Galton–Watson trees.) *Let T be a super-critical Bienaymé–Galton–Watson tree whose offspring distribution has mean $m > 1$ and finite second moment, and is conditioned on non-extinction.*

(*i*) *Let $(\ell_n)_{n\in\mathbb{N}}$ be a sequence of integers such that $0 \le \ell_n \le n$ for all $n \in \mathbb{N}$. For every $n \in \mathbb{N}$, let $A_n = \cup_{k=(n-\ell_n)_+}^{n} G_k$ be the subset composed of the last $\ell_n + 1$ generations of T_n. Then, the sequence $(A_n)_{n\in\mathbb{N}}$ a.s. satisfies Assumptions 1 and 2.*

(*ii*) *The sequences $(G_n)_{n\in\mathbb{N}}$ and $(T_n)_{n\in\mathbb{N}}$ a.s. satisfy Assumptions 1 and 2.*

In the case of a critical Bienaymé–Galton–Watson tree it is not possible to condition on non-extinction, but taking $A_n = G_n$ and conditioning on non-extinction at time n, [2, Theorem 2.1] suggests that Assumption 1 should be satisfied but not Assumption 2.

The proof of Lemma 7 (i) relies on the case where the sequence $(\ell_n)_{n\in\mathbb{N}}$ is bounded. In this case, we prove a stronger version of [3, Theorem 2] where individuals can have zero children and the two random individuals are taken from the last $\ell_n + 1$ generations instead of only the last generation.

Before proving Lemma 7, we need to prove the following lemma stating that the last generations carry most of the weight of T_n.

Lemma 8. (Last generations carry all the weight.) *Let T be a super-critical Bienaymé–Galton–Watson tree whose offspring distribution has mean $m > 1$ and is conditioned on non-extinction. Then,*

$$\text{for all } \ell \in \mathbb{N},\ a.s.,\quad \lim_{n\to\infty} \frac{|\cup_{k=(n-\ell)_+}^{n} G_k|}{|T_n|} = 1 - \frac{1}{m^{\ell+1}}. \tag{6}$$

Proof. For all $n \in \mathbb{N}$, we write $Z_n = |G_n|$. Using [6, Theorem I.C.12.3], as T is a super-critical Bienaymé–Galton–Watson tree whose offspring distribution $\mathcal{P}$ has a finite mean $m > 1$ and is conditioned on non-extinction, we know that there exists a sequence of positive constants $(C_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} C_n = \infty$ and $\lim_{n\to\infty} C_{n+1}/C_n = m$, and such that, a.s., $\lim_{n\to\infty} C_n^{-1} Z_n = W$, where W is a random variable on $\mathbb{R}_+$, and where the event $\{W = 0\} \supset \{\text{there exist } n,\ Z_n = 0\}$ has zero probability. Hence, a.s., $C_n^{-1} |\cup_{k=0}^{\ell} G_{n-k}| = C_n^{-1} \sum_{k=0}^{\ell} Z_{n-k}$ converges to $W \sum_{k=0}^{\ell} m^{-k}$ as n goes to infinity. Writing $S_n = \sum_{k=0}^{n} C_k$, as we know that $C_n \sim mC_{n-1}$, we have $S_{n+1} = S_n + C_{n+1} = mS_n + o(S_n)$, and thus $C_n \sim C_{n+1}/m \sim ((m-1)/m)S_n$. As $\lim_{n\to\infty} C_n^{-1} \sum_{k=0}^{n} C_k = m/(m-1)$, using the Stolz–Cesàro theorem, we get that, a.s., $C_n^{-1}|T_n| = C_n^{-1} \sum_{k=0}^{n} Z_k$ converges to $Wm/(m-1)$ as n goes to infinity. Consequently, we get (6). □

Proof of Lemma 7. Part (ii) is an immediate consequence of part (i), taking $\ell_n = 0$ or n.

We now prove part (i). Using Lemma 5 with (6), it is enough to prove that the sequence $(A_n)_{n\in\mathbb{N}}$ a.s. satisfies Assumption 2. Let U_n and V_n be independent and uniformly sampled elements of A_n. For $k \in \mathbb{N}$, note that $h(U_n \wedge V_n) \ge k$ is equivalent to the existence of $u \in G_k$ such that $U_n, V_n \in T(u)$. Note that Assumption 2 for $(A_n)_{n\in\mathbb{N}}$ in the random tree T can be reformulated as

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}(h(U_n \wedge V_n) \ge k \mid T) = \lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}(\text{there exists } u \in G_k,\ U_n, V_n \in T(u) \mid T) = 0. \tag{7}$$

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press

We divide the rest of the proof into two cases: we first consider the case when the sequence $(\ell_n)_{n\in\mathbb{N}}$ is bounded, and then the general case.

Case 1: The sequence $(\ell_n)_{n\in\mathbb{N}}$ is bounded. Following [6, Section I.D.12] (and using the survivor/extinct vertices denomination from [1, Section 2.2.3]), the vertices of the super-critical Bienaymé–Galton–Watson tree T conditioned on non-extinction can be partitioned into two categories: *survivor* vertices, whose descendants do not suffer extinction, and *extinct* vertices, whose descendants eventually become extinct. The root of T is a survivor vertex due to the conditioning on non-extinction. Denote by T^{s} the subtree of T composed of survivor vertices. From [6, Theorem I.D.12.1], the tree T^{s} is distributed as a super-critical Bienaymé–Galton–Watson tree whose offspring distribution $\mathcal{P}_0$ on $\mathbb{N}^*$ (hence no extinction) has the same mean $m>1$ as $\mathcal{P}$ and has finite second moment. Also, from [6, Theorem I.D.12.3], we know that if $u\in T$ is an extinct vertex, then its descendant subtree $T(u)$ is distributed as a sub-critical Bienaymé–Galton–Watson tree whose offspring distribution $\mathcal{P}_1$ is explicitly known and has finite second moment.

For all $n>\ell$ and $u\in T$, define $Z_{n,\ell,u}=|(T_n\setminus T_{n-\ell-1})\cap T(u)|$. Fix some $k\in\mathbb{N}$. Conditionally on T, the unique ancestor of U_n in G_k is $u\in G_k$ with probability $Z_{n,\ell_n,u}/\sum_{v\in G_k} Z_{n,\ell_n,v}$. Define $Z_n=|G_n|$, $G_n^{\mathrm{s}}=G_n\cap T^{\mathrm{s}}$, and $Z_n^{\mathrm{s}}=|G_n^{\mathrm{s}}|$. Then

$$\mathbb{P}(\text{there exists } u\in G_k,\ U_n,\ V_n\in T(u)\mid T)=\frac{\sum_{u\in G_k} Z_{n,\ell_n,u}^2}{\big(\sum_{u\in G_k} Z_{n,\ell_n,u}\big)^2}=\frac{\sum_{u\in G_k^{\mathrm{s}}} Z_{n,\ell_n,u}^2}{\big(\sum_{u\in G_k^{\mathrm{s}}} Z_{n,\ell_n,u}\big)^2},$$

where the last equality holds a.s. for n large enough since, for any extinct vertex $u\in G_k\setminus G_k^{\mathrm{s}}$, we know that, a.s., $Z_{n,\ell_n,u}=0$ for n large enough (recall that the sequence $(\ell_n)_{n\in\mathbb{N}}$ is bounded).

From [6, Theorems I.B.6.1 and I.B.6.2], as $\mathcal{P}$ has finite second moment, we know that $\lim_{n\to\infty} m^{-n}Z_n=W$ a.s. and in L^2, where W is a random variable in $\mathbb{R}_+^*$ with finite second moment. (From [6, Theorem I.B.6.2-(iii)], we know that on the non-extinction event, that is, when the root is a survivor vertex, a.s. W is positive.) Then, note that, for all survivor vertices $u\in G_k^{\mathrm{s}}$, we have, a.s., $\lim_{n\to\infty} m^{-(n-k)}Z_{n,0,u}=W_u$, where, conditionally on G_k^{s}, the random variables $(W_u)_{u\in G_k^{\mathrm{s}}}$ are independent and distributed as W (recall that the root of T is also a survivor vertex). Thus we get, a.s. for all $\ell\in\mathbb{N}$ and all $u\in G_k^{\mathrm{s}}$, $\lim_{n\to\infty} m^{-(n-k)}Z_{n,\ell,u}=\big(\sum_{j=0}^{\ell} m^{-j}\big)W_u$. As the sequence $(\ell_n)_{n\in\mathbb{N}}$ is bounded, this implies that, a.s. for all vertices $u\in G_k^{\mathrm{s}}$, $\lim_{n\to\infty} m^{-(n-k)}\big(\sum_{j=0}^{\ell_n} m^{-j}\big)^{-1} Z_{n,\ell_n,u}=W_u$. Hence, almost surely,

$$\frac{\sum_{u\in G_k^{\mathrm{s}}} Z_{n,\ell,u}^2}{\big(\sum_{u\in G_k^{\mathrm{s}}} Z_{n,\ell,u}\big)^2}\xrightarrow[n\to\infty]{}\frac{\sum_{u\in G_k^{\mathrm{s}}} W_u^2}{\big(\sum_{u\in G_k^{\mathrm{s}}} W_u\big)^2}.$$

Combining what we have so far, we get

$$\text{almost surely,}\qquad \lim_{n\to\infty}\mathbb{P}(h(U_n\wedge V_n)\geq k\mid T)=\frac{\sum_{u\in G_k^{\mathrm{s}}} W_u^2}{\big(\sum_{u\in G_k^{\mathrm{s}}} W_u\big)^2}. \tag{8}$$

As the left-hand side in (8) is a non-increasing function of k, then so is the right-hand side in (8). Thus, taking the limit when $k\to\infty$ and then taking the expectation, we get

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press

$$\mathbb{E}\Big[\lim_{k\to\infty}\lim_{n\to\infty}\mathbb{P}(h(U_n\wedge V_n)\geq k\mid T)\Big]=\mathbb{E}\left[\lim_{k\to\infty}\frac{\sum_{u\in G_k^{\mathrm{s}}}W_u^2}{\big(\sum_{u\in G_k^{\mathrm{s}}}W_u\big)^2}\right]$$
$$=\lim_{k\to\infty}\mathbb{E}\left[\frac{\sum_{u\in G_k^{\mathrm{s}}}W_u^2}{\big(\sum_{u\in G_k^{\mathrm{s}}}W_u\big)^2}\right],\tag{9}$$

where we used the dominated convergence theorem in the last inequality.

Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables distributed as W. For all $n\in\mathbb{N}$, define the random variable $R(n)=\sum_{i=1}^n W_i^2/\big(\sum_{i=1}^n W_i\big)^2$. Using the strong law of large numbers, we get that the numerator of $R(n)$ is a.s. equivalent to $n\mathbb{E}[W^2]$, and the denominator of $R(n)$ is a.s. equivalent to $n^2\mathbb{E}[W]^2$. This implies that $R(n)$ is a.s. equivalent to $n^{-1}\mathbb{E}[W^2]/\mathbb{E}[W]^2$, and thus, a.s., $\lim_{n\to\infty}R(n)=0$. Note that the expectation in the last line of (9) can be written as $\mathbb{E}[R(Z_k^{\mathrm{s}})]$. As we know that, a.s., $\lim_{k\to\infty}Z_k^{\mathrm{s}}=\infty$, we get that, a.s., $\lim_{k\to\infty}R(Z_k^{\mathrm{s}})=0$. Hence, using the dominated convergence theorem (with domination by 1), we get that $\lim_{n\to\infty}\mathbb{E}[R(Z_k^{\mathrm{s}})]=0$, which implies that the left-hand side in (9) is also null. As a consequence, a.s. (7) holds, which implies that, a.s., the sequence $(A_n)_{n\in\mathbb{N}}$ satisfies Assumption 2. This concludes the proof of the first case.

Case 2: General case. Let $\varepsilon>0$, and let $\ell\in\mathbb{N}$ be such that $m^{-\ell-1}<\varepsilon$. Then, using Lemma 8, we get that, a.s., $\mathbb{P}(h(U_n)<n-\ell\mid T)<\varepsilon$ for n large enough. Thus, a.s. for n large enough, for all $k\in\mathbb{N}$,

$$\begin{aligned}\mathbb{P}(h(U_n\wedge V_n)\geq k\mid T)&\leq\mathbb{P}(h(U_n)<n-\ell\text{ or }h(V_n)<n-\ell\mid T)\\&\quad+\mathbb{P}(h(U_n)\geq n-\ell\text{ and }h(V_n)\geq n-\ell\mid T)\\&\quad\times\mathbb{P}(h(U_n\wedge V_n)\geq k\mid\min\left(h(U_n),h(V_n)\right)\geq n-\ell,\ T)\\&\leq 2\varepsilon+\mathbb{P}(h(\bar U_n\wedge\bar V_n)\geq k\mid T),\end{aligned}\tag{10}$$

where $\bar U_n$ and $\bar V_n$ are independent and uniformly distributed over $\cup_{k=(n-\min(\ell_n,\ell))_+}^n G_k$. By the first case, $\lim_{k\to\infty}\limsup_{n\to\infty}\mathbb{P}(h(\bar U_n\wedge\bar V_n)\geq k\mid T)=0$. Thus, using (10), we get $\lim_{k\to\infty}\limsup_{n\to\infty}\mathbb{P}(h(U_n\wedge V_n)\geq k\mid T)\leq 2\varepsilon$. This being true for all $\varepsilon>0$, we get that, a.s., the sequence $(A_n)_{n\in\mathbb{N}}$ satisfies Assumption 2, which concludes the proof. □

4. Dependence of the variance on the shape of the tree

In this section, we briefly discuss the variance of the empirical average estimator $\bar M_A(f)$ (which estimates $\langle\mu,f\rangle$), i.e. $\mathbb{E}_\mu[|A|^{-1}M_A(f)^2]$ for some Borel function f on $\mathcal{X}$, and its dependence on the geometry of the averaging set $A\subset T$. We consider the case where the transition kernel Q induces a self-adjoint compact operator on $L^2(\mu)$, where μ is the unique invariant measure of Q (note that self-adjoint is equivalent to (μ,Q) being reversible). This is, in particular, the case when the state space $\mathcal{X}$ is finite and (μ,Q) is reversible, or when the operator induced by Q on $L^2(\mu)$ is a symmetric Hilbert–Schmidt operator.

We now prove Proposition 1, which is a non-asymptotic result that states that, among subtrees $T\subset\mathbb{T}^\infty$ of a given finite size, the line graph tree (i.e. the Markov chain) is the one minimizing the variance of the empirical average estimator.

Proof of Proposition 1. Let $f\in L^2(\mu)$ be some function. As Q induces a self-adjoint compact operator on $L^2(\mu)$, using [11, Theorem 12.29(d) and Theorem 12.30] (recall that a self-adjoint

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press

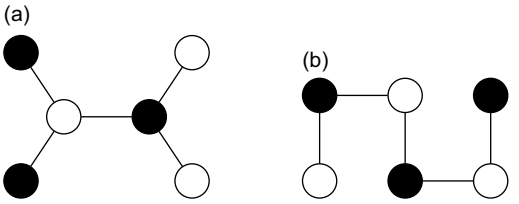


FIGURE 1. Comparison of the double-cherry graph and the line graph ($n = 6$); both graphs have an exactly balanced bipartite 2-coloring, and thus satisfy $H_T(-1) = 0$.

operator is normal), the spectrum of Q is composed of (at most) a countable number of eigenvalues $(\alpha_k)_{k\in\mathbb{N}}$, and the function $f \in L^2(\mu)$ has a unique expansion $\sum_{k\in\mathbb{N}} f_k$ where $Qf_k = \alpha_k f_k$ for all $k \in \mathbb{N}$, and $\langle f_k, f_\ell\rangle_{L^2(\mu)} = 0$ for $k \neq \ell$. As Q is a Markov kernel, we have $\alpha_k \in [-1, 1]$ for all $k \in \mathbb{N}$ (indeed, using Jensen's inequality, we have $\alpha_k^2\langle \mu, f_k^2\rangle = \langle \mu, (Qf_k)^2\rangle \le \langle \mu, Q(f_k^2)\rangle = \langle \mu, f_k^2\rangle$). Note that, for all $n, m \in \mathbb{N}$,

$$\langle \mu, Q^n f \times Q^m f\rangle = \sum_{i,j\in\mathbb{N}} \langle Q^n f_i, Q^m f_j\rangle_{L^2(\mu)} = \sum_{i\in\mathbb{N}} \alpha_i^{n+m}\langle \mu, f_i^2\rangle.$$

Using (2) with $\nu = \mu$, we get that $\mathbb{E}_\mu[f(X_u)f(X_v)] = \sum_{k\in\mathbb{N}} \alpha_k^{d(u,v)}\langle \mu, f_k^2\rangle$. In particular, $\mathbb{E}_\mu[M_A(f)^2] = \sum_{k\in\mathbb{N}} \mathbb{E}_\mu[M_A(f_k)^2]$, and thus it is sufficient to prove the result when $f = f_k$ for some $k \in \mathbb{N}$. (Note that there exists $k \in \mathbb{N}$ such that $f_k \neq 0$ and $\alpha_k \notin \{-1, 0, 1\}$ if and only if $f \notin \mathrm{Ker}(Q) \oplus \mathrm{Ker}(Q - I) \oplus \mathrm{Ker}(Q + I)$.)

Hence, let $k \in \mathbb{N}$ be fixed, and in the rest of the proof we will write $f = f_k$ and $\alpha = \alpha_k \in [-1, 1]$. Thus, the variance of the empirical average estimator can be written as

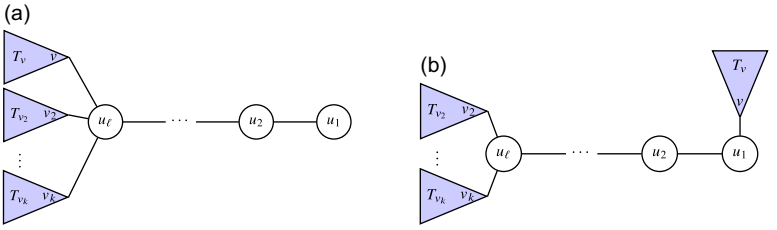
$$|A|^{-1}\,\mathbb{E}_\mu[M_A(f)^2] = |A|^{-1}\langle \mu, f^2\rangle\, H_A(\alpha), \quad \text{with } H_A(\alpha) = \sum_{u,v\in A} \alpha^{d(u,v)},$$

which involves the Hosoya–Wiener polynomial $H_A(\alpha)$ of the subset A. Note that variance minimization is equivalent to minimization of $H_A(\alpha)$. Also note that we may consider unrooted trees $A = T$, as the definition of $H_T(\alpha)$ is invariant by rerooting the tree. Hence, the proof of the proposition is completed by applying Lemma 1. □

We now prove Lemma 1, which states that the Hosoya–Wiener polynomial is minimized by the line graph tree among trees of a given size.

Proof of Lemma 1. If $\alpha = 0$ or 1, then the value of the Hosoya–Wiener polynomial H_T depends only on the size of T, and thus is the same for every tree T of size n. If $\alpha = -1$, then we have $H_T(-1) = (|B| - |R|)^2$, where $T = B \cup R$ is the bipartite partitioning of vertices in T, i.e. the value of $H_T(-1)$ is the imbalance between the two bipartite classes of vertices in T (2-coloring of T), and its minimal value is 0 (resp. 1) when n is even (resp. odd), and is achieved by the line graph (but not uniquely for $n \ge 5$, e.g. the double-cherry graph in Figure 1a is also a minimizer for $n = 6$). We now assume that $\alpha \in (-1, 1) \setminus \{0\}$, and we are going to prove that in this case the line graph is the unique minimizer of $H_T(\alpha)$. We divide the rest of the proof into two cases depending on the sign of α.

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press



The unrooted tree T before modification

The unrooted tree T' before modification

FIGURE 2. The unrooted trees T and T' for $\alpha \in (0, 1)$.

Proof for $\alpha \in (0, 1)$. The case $\alpha \in (0, 1)$ was already proved in [8, Theorem 9], but as it is a simple argument, we include it here for the sake of completeness. Let $n \in \mathbb{N}^*$. To prove that the line graph minimizes the function $T \mapsto H_T(\alpha)$ among trees of size n, let u_1 be a leaf of T, and consider the tree T to be rooted at u_1. Let $\ell \in \mathbb{N}^*$ be the first generation with size larger than 1 (which exists as T is not the line graph), and, for $i < \ell$, denote by u_{i+1} the only vertex in the ith generation. Let v be one of the children of the vertex u_ℓ, and denote by T_v the connected component of $T \setminus \{u_\ell\}$ that contains v. We define the tree T' by removing the edge (u_ℓ, v) and adding the edge (u_1, v) (see Figure 2b, where for clarity the other children of u_ℓ have been labeled $v_2, \dots, v_k$). We now compare the distances $d_{T'}(u, w)$ and $d_T(u, w)$ for $u, w \in T$: if u and w are both in $T \setminus T_v$ or both in T_v, then $d_{T'}(u, w) = d_T(u, w)$; if $w \in T_v$, then $d_{T'}(u_i, w) = d_T(u_{\ell-i+1}, w)$ for $1 \le i \le \ell$; and if $u \in T \setminus (T_v \cup \{u_1, \dots, u_\ell\})$ and $w \in T_v$, then

$$d_{T'}(u, w) = d_{T'}(u, u_\ell) + d_{T'}(u_\ell, v) + d_{T'}(w, v) = d_T(u, u_\ell) + \ell + d_T(w, v) > d_T(u, w).$$

Hence, $H_T(\alpha) - H_{T'}(\alpha) = 2 \sum_{u \in T_v, w \in T \setminus (T_v \cup \{u_1, \dots, u_\ell\})} \alpha^{d_T(u,v)} - \alpha^{d_{T'}(u,v)} > 0$, and thus T does not minimize $H_T(\alpha)$.

Proof for $\alpha \in (-1, 0)$. We prove the statement by recurrence on the size n of the tree. For $n \in \{1, 2, 3\}$, there exists only one (unrooted) tree of size n, hence the statement is trivial.

Before proving that this property is hereditary, first note that if T is a tree of size $n \ge 2$ and $u, v \in T$ are two vertices connected by an edge, and we let T_u and T_v be the two rooted subtrees of T obtained by removing the edge (u,v) and rooted at u and v respectively, then we get

$$H_T(\alpha) = H_{T_u}(\alpha) + H_{T_v}(\alpha) + 2 \sum_{u' \in T_u} \sum_{v' \in T_v} \alpha^{d_T(u', v')} = H_{T_u}(\alpha) + H_{T_v}(\alpha) + 2\alpha\, C_{T_u}(\alpha) C_{T_v}(\alpha), \tag{11}$$

where for convenience we write $C_{T_v}(\alpha) = \left(\sum_{v' \in T_v} \alpha^{d_T(v, v')} \right)$, and similarly for T_u. We also remark that when T_v is the line graph with k vertices $\{u_1, \dots, u_k\}$ rooted at the vertex $v = u_j$, then we have $C_{T_v}(\alpha) = (1 + \alpha - \alpha^j - \alpha^{k-j+1})/(1 - \alpha)$, which is maximal among rooted copies of $\{u_1, \dots, u_k\}$ only for $j = 1$ or k (recall that $\alpha \in (-1, 0)$). We denote by L_k the line graph with k vertices rooted at one of its extremities.

Now, consider $n > 3$ and assume that, for all $k < n$, the line graph is the unique minimizer of $T \mapsto H_T(\alpha)$ among trees of size k. Consider an unrooted tree T of size n that is not the line graph. Let u be a leaf of the tree obtained by removing all the leaves of T (such a vertex is sometimes called a non-protected vertex). In particular, the node u has $2 + \ell$ neighbors in T with $\ell \in \mathbb{N}$, which we denote by $u_0, \dots, u_{\ell+1}$; and at most one of the neighbors of u is not a leaf, say $v = u_{\ell+1}$. (All the neighbors of u are leaves if and only if T is the star graph

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press

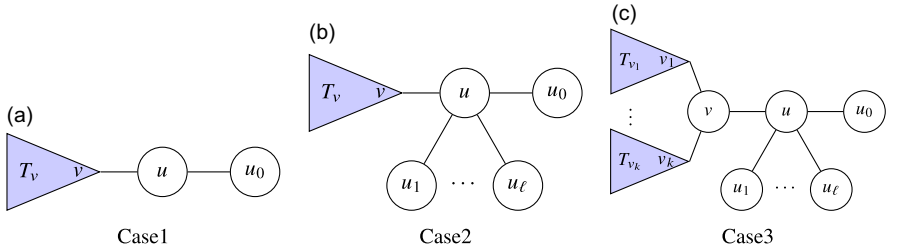


FIGURE 3. The unrooted tree T before modification.

whose center vertex is u, in which case we still write $v = u_{\ell+1}$.) Denote by T_v the subtree $T \setminus \{u, u_0, \ldots, u_\ell\}$ of T rooted at v. We consider three cases.

Case 1: $\ell = 0$ (*see* Figure 3a). As T is not the line graph and as T_v has size $n-2$, by the induction hypothesis we either have $H_{T_v}(\alpha) > H_{L_{n-2}}(\alpha)$, or T_v is a line graph rooted at a non-extremal vertex (and thus $C_{T_v}(\alpha) < C_{L_{n-2}}(\alpha)$). If $C_{T_v}(\alpha) < C_{L_{n-2}}(\alpha)$, using (11) around edge (v,u), as $\alpha(1+\alpha) < 0$ (recall that $\alpha \in (-1, 0)$), we get

$$\begin{aligned} H_T(\alpha) &= H_{T_v}(\alpha) + H_{L_2}(\alpha) + 2\alpha C_{T_v}(\alpha)(1+\alpha) \\ &> H_{L_{n-2}}(\alpha) + H_{L_2}(\alpha) + 2\alpha C_{L_{n-2}}(\alpha)(1+\alpha) = H_{L_n}(\alpha). \end{aligned}$$

If $C_{T_v}(\alpha) \geq C_{L_{n-2}}(\alpha)$ (thus we have $H_{T_v}(\alpha) > H_{L_{n-2}}(\alpha)$), using (11) around edge (u, u_0), since $C_{L_{n-1}}(\alpha) = 1 + \alpha C_{L_{n-2}}(\alpha)$,

$$\begin{aligned} H_T(\alpha) &= H_{T\setminus\{u_0\}}(\alpha) + H_{L_1}(\alpha) + 2\alpha(1 + \alpha C_{T_v}(\alpha)) \\ &> H_{L_{n-1}}(\alpha) + H_{L_1}(\alpha) + 2\alpha(1 + \alpha C_{L_{n-2}}(\alpha)) = H_{L_n}(\alpha). \end{aligned}$$

Thus, for all values of $C_{T_v}(\alpha)$, we get that $H_T(\alpha) > H_{L_n}(\alpha)$.

Case 2: $\ell \geq 1$ *and* $C_{T_v}(\alpha) > 0$. Denote by T_u the subtree of T composed of the $\ell + 2$ vertices $u, u_0, \ldots, u_\ell$ (see Figure 3b), and consider the tree T' obtained from T by replacing T_u by a copy of $L_{2+\ell}$. By the induction hypothesis, we know that $H_{T_u}(\alpha) \geq H_{L_{2+\ell}}(\alpha)$. Using (11) around edge (v,u), since $1 + (\ell+1)\alpha < C_{L_{2+\ell}}(\alpha)$,

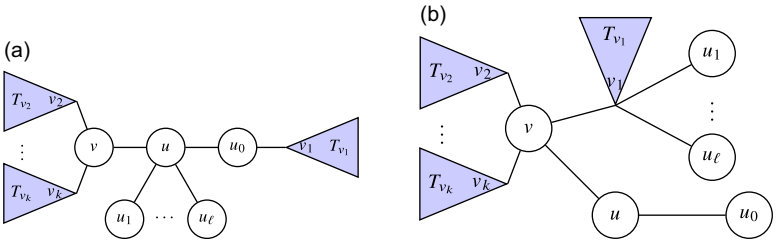
$$\begin{aligned} H_T(\alpha) &= H_{T_v}(\alpha) + H_{T_u}(\alpha) + 2\alpha C_{T_v}(\alpha)(1 + (\ell+1)\alpha) \\ &> H_{T_v}(\alpha) + H_{L_{2+\ell}}(\alpha) + 2\alpha C_{T_v}(\alpha) C_{L_{2+\ell}}(\alpha) = H_{T'}(\alpha), \end{aligned}$$

and thus the tree T does not minimize the Hosoya–Wiener polynomial.

Case 3: $\ell \geq 1$ *and* $C_{T_v}(\alpha) \leq 0$. As $C_{T_v}(\alpha) \neq 1$, we know that v is not a leaf. Denote by $v_1, \ldots, v_k$ the neighbors of v other than u, and for each $i \in [\![1, k]\!]$, denote by T_{v_i} the subtree of T_v rooted at v_i (see Figure 3c). As $C_{T_v}(\alpha) = 1 + \alpha \sum_{i=1}^{k} C_{T_{v_i}}(\alpha) \leq 0$, we have $\sum_{i=1}^{k} C_{T_{v_i}}(\alpha) \geq -1/\alpha > 0$, and thus there exists $i \in [\![1, k]\!]$ such that $C_{T_{v_i}}(\alpha) > 0$. Without loss of generality, we assume for simplicity that $i = 1$.

Case 3a: $\ell \geq 1$ *and* $C_{T_v}(\alpha) \leq 0$ *and* $\sum_{i=2}^{k} C_{T_{v_i}}(\alpha) > 0$. Consider the tree T' obtained from T by replacing edge (v_1, v) by (v_1, u_0), i.e. grafting T_{v_1} on u_0 (see Figure 4a). Note that going from T to T', the subtrees T_{v_1} and $T \setminus T_{v_1}$ do not change, the distance between v_1 and v_i for $i \in [\![2, k]\!]$ goes from 2 to 4, and v_1 is still at distance 1 from a leaf of the

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press



Case 3a after moving T_{v_1} Case 3b after moving $u_1, \dots, u_\ell$

FIGURE 4. The unrooted tree T' in Case 3 after modification.

star graph formed by the vertices $\{v, u, u_0, \dots, u_\ell\}$. Thus, as $\alpha \in (-1, 0)$,

$$H_T(\alpha) - H_{T'}(\alpha) = 2C_{T_{v_1}}(\alpha)\left(\sum_{i=2}^{k} C_{T_{v_i}}(\alpha)\right)(\alpha^2 - \alpha^4) > 0,$$

and thus the tree T does not minimize the Hosoya–Wiener polynomial.

Case 3b: $\ell \geq 1$ *and* $C_{T_v}(\alpha) \leq 0$ *and* $\sum_{i=2}^{k} C_{T_{v_i}}(\alpha) \leq 0$. In particular, we get that $C_{T_{v_1}}(\alpha) \geq -1/\alpha \geq 1$. Consider the tree T' obtained from T by replacing edge (u_i, u) by (u_i, v_1) for all $i \in [\![1, \ell]\!]$, i.e. grafting leaves $u_1, \dots, u_\ell$ to v_1 (see Figure 4b). Note that going from T to T', the subtree $T \setminus \{u_1, \dots, u_\ell\}$ does not change, and for $i, j \in [\![1, \ell]\!]$, the distance between u_i and u_j (resp. v) does not change, and the distance between u_i and v_1 (resp. u) goes from 3 to 1 (resp. from 1 to 3). Thus, as $\alpha \in (-1, 0)$,

$$\begin{aligned} H_T(\alpha) - H_{T'}(\alpha) &= 2\ell C_{T_{v_1}}(\alpha)(\alpha^3 - \alpha) + 2\ell(1+\alpha)(\alpha - \alpha^3) \\ &\geq 2\ell(\alpha^3 - \alpha) + 2\ell(1+\alpha)(\alpha - \alpha^3) \geq 2\ell(\alpha^2 - \alpha^4) > 0, \end{aligned}$$

and thus the tree T does not minimize the Hosoya–Wiener polynomial. □

Funding information

There are no funding bodies to thank relating to the creation of this article.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] ABRAHAM, R. AND DELMAS, J.-F. (2020). An introduction to Galton–Watson trees and their local limits. Preprint, arXiv:1506.05571.
[2] ATHREYA, K. B. (2012). Coalescence in critical and subcritical Galton–Watson branching. *J. Appl. Prob.* **49**, 627–638.
[3] ATHREYA, K. B. (2012). Coalescence in the recent past in rapidly growing populations. *Stoch. Process. Appl.* **122**, 3757–3766.
[4] ATHREYA, K. B. AND KANG, H.-J. (1998). Some limit theorems for positive recurrent branching Markov chains: I. *Adv. Appl. Prob.* **30**, 693–710.
[5] ATHREYA, K. B. AND KANG, H.-J. (1998). Some limit theorems for positive recurrent branching Markov chains: II. *Adv. Appl. Prob.* **30**, 711–722.

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press

[6] ATHREYA, K. B. AND NEY, P. E. (1972). *Branching Processes*, Springer, Berlin.
[7] BANSAYE, V. (2019). Ancestral lineages and limit theorems for branching Markov chains in varying environment. *J. Theoret. Prob.* **32**, 249–281.
[8] CASABLANCA, R. M. AND DANKELMANN, P. (2019). Distance and eccentric sequences to bound the Wiener index, Hosoya polynomial and the average eccentricity in the strong products of graphs. *Discrete Appl. Math.* **263**, 105–117.
[9] DELMAS, J.-F. AND MARSALLE, L. (2010). Detection of cellular aging in a Galton–Watson process. *Stoch. Process. Appl.* **120**, 2495–2519.
[10] GUYON, J. (2007). Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Prob.* **17**, 1538–1569.
[11] RUDIN, W. (1996). *Functional Analysis*, 2nd edn. McGraw-Hill, Boston, MA.
[12] TIAN, D. AND CHOI, K. P. (2013). Sharp bounds and normalization of Wiener-type indices. *PLoS ONE* **8**, e78448.
[13] WAGNER, S., WANG, H. AND ZHANG, X.-D. (2013). Distance-based graph invariants of trees and the Harary index, *Filomat* **1**, 41–50.

https://doi.org/10.1017/jpr.2025.14 Published online by Cambridge University Press