Probab. Theory Relat. Fields 116, 41–56 (2000)

Steven P. Lalley · Thomas Sellke

An extension of Hawkes' theorem on the Hausdorff dimension of a Galton–Watson tree

Received: 30 June 1998 / Revised: 4 February 1999

Abstract. Let $\mathscr{T}$ be the genealogical tree of a supercritical multitype Galton–Watson process, and let Λ be the *limit set* of $\mathscr{T}$, i.e., the set of all infinite self-avoiding paths (called *ends*) through $\mathscr{T}$ that begin at a vertex of the first generation. The limit set Λ is endowed with the metric $d(\zeta,\xi)=2^{-n}$ where $n=n(\zeta,\xi)$ is the index of the first generation where ζ and ξ differ. To each end ζ is associated the infinite sequence $\Phi(\zeta)$ of types of the vertices of ζ. Let Ω be the space of all such sequences. For any ergodic, shift-invariant probability measure μ on Ω, define Ω_μ to be the set of all μ*-generic* sequences, i.e., the set of all sequences $\omega\in\Omega$ such that each finite sequence v occurs in ω with limiting frequency $\mu(\Omega(v))$, where $\Omega(v)$ is the set of all $\omega'\in\Omega$ that begin with the word v. Then the Hausdorff dimension of $\Lambda\cap\Phi^{-1}(\Omega_\mu)$ in the metric d is

$$(h(\mu)+\int_\Omega \log q(\omega_0,\omega_1)d\mu(\omega))_+/\log 2\ ,$$

almost surely on the event of nonextinction, where $h(\mu)$ is the entropy of the measure μ and $q(i,j)$ is the mean number of type-j offspring of a type-i individual. This extends a theorem of HAWKES [5], which shows that the Hausdorff dimension of the entire boundary at infinity is $\log_2\alpha$, where α is the Malthusian parameter.

1. Introduction

Suppose that the individuals of a simple supercritical Galton–Watson process are randomly assigned labels from a finite set $\mathscr{L}$. On the event of nonextinction, there will be infinite *descent lines* (sequences $\xi=\xi_1\xi_2\cdots$ of individuals such that each ξ_{n+1} is an offspring of ξ_n, and such that ξ_1 is a member of the first generation). Define the *limit set* Λ of the Galton–Watson process to be the set of all infinite descent lines ξ; equivalently, Λ is the *space of ends* of the genealogical tree of the Galton-Watson process. For each infinite line of descent $\xi\in\Lambda$, define its *pedigree* $\Phi(\xi)=l_1l_2\cdots$ to be the infinite sequence of labels asssigned to the individuals $\xi_1,\xi_2,\ldots$ of ξ. A subset B of $\Omega=\mathscr{L}^{\mathbb{N}}$ is called *non-polar* if there is positive probability that $B\cap\Phi(\Lambda)\neq\emptyset$. Characterizations of non-polar sets B have been given in [3, 8, 9].

S.P. Lalley: University of Chicago, Department of Statistics, Chicago, IL 60637, USA. e-mail: `lalley@stat.purdue.edu`

T. Sellke: Purdue University, Department of Statistics, West Lafayette, IN 47907, USA. e-mail: `tsellke@stat.purdue.edu`

Supported by NSF Grant DMS-9626590.

Mathematics Subject Classification (1991): 60J80

This paper concerns the *size* of $\Lambda \cap \Phi^{-1}(B)$ for certain non-polar sets B. The sets of primary interest are the sets $\Omega_\mu \subset \Omega$ consisting of all μ*-generic* sequences, where μ is an ergodic, shift-invariant probability measure on Ω. A sequence $\omega \in \Omega$ is called μ-generic if every finite sequence $x = x_1 x_2 \cdots x_n$ with entries in $\mathscr{L}$ occurs with limiting frequency $\mu(\Omega(x))$ in ω, where $\Omega(x)$ is the set of all $\omega' \in \Omega$ whose first n entries coincide with the entries x_i of x. The natural way to measure the size of $\Lambda \cap \Phi^{-1}(\Omega_\mu)$ is by *Hausdorff dimension*. The idea of using Hausdorff dimension as a measure of size for a Galton–Watson tree (or its space of ends) derives from a paper of HAWKES [5] (see also [4] and [8]). Let $d = d_\Lambda$ and $d = d_\Omega$ be the metrics on Λ and Ω defined by

$$d(\xi, \zeta)) = 2^{-n(\xi,\zeta)} \quad , \tag{1}$$

respectively, where $n(\xi, \zeta)$ is the smallest integer n such that $\xi_n \neq \zeta_n$. Hawkes proved the following theorem:

Hawkes' theorem. *If the offspring distribution has mean $\alpha > 1$ and finite second moment then, almost surely on the event of nonextinction, the limit set Λ of the Galton–Watson tree has Hausdorff dimension* (*in the metric d_Λ*)

$$\delta_H(\Lambda) = \frac{\log \alpha}{\log 2} \quad . \tag{2}$$

R. Lyons [8] subsequently showed that the second moment hypothesis is unecessary.

The main result of this paper is an explicit formula for the Hausdorff dimension of $\Lambda \cap \Phi^{-1}(\Omega_\mu)$ in the metric d defined by (1). Apart from its intrinsic interest, this result is crucial for the main results of [6] concerning the "backscattering" phenomenon for anisotropic branching random walks on homogeneous trees. The formula holds generally for a *multi-type* Galton–Watson process $\mathscr{G}$ (see [1], Ch. V for background) with finite type space $\mathscr{L}$. For each pair $i, j \in \mathscr{L}$ let $q_{ij} = q(i, j)$ be the mean number of type-j offspring produced by a type-i individual, and let $Q = (q_{ij})_{i,j\in\mathscr{L}}$ be the $|\mathscr{L}| \times |\mathscr{L}|$ matrix of means. We shall make the following assumptions:

(H1) *The mean offspring numbers q_{ij} are all finite.*
(H2) *The matrix Q is irreducible*; *equivalently*, *some positive power Q^n of Q has strictly positive entries.*

For simplicity, we shall assume that the 0th generation consists of a single individual of type i_*, where i_* is a distinguished element of $\mathscr{L}$. Thus, the genealogical tree $\mathscr{T}$ has a single root vertex labelled i_*. The vertices $\mathscr{V}$ of $\mathscr{T}$ may be partitioned by *depth*:

$$\mathscr{V} = \cup_{n=0}^{\infty} \mathscr{V}_n \quad ,$$

where the elements of $\mathscr{V}_n$ are the nth generation individuals of the underlying Galton–Watson process. By (H1), all of the generations $\mathscr{V}_n$ are finite, and by convention $\mathscr{V}_0$ is a singleton containing the root vertex x_0. The (directed) edges of $\mathscr{T}$ connect those ordered pairs (v, w) of vertices such that w is an offspring of v. If there is a directed path starting at a vertex v and ending at a vertex w then w is called a *descendant* of v; the subtree of $\mathscr{T}$ whose vertices are the descendants of v

is denoted by $\mathscr{T}(v)$. An infinite (connected) path in $\mathscr{T}$ starting in $\mathscr{V}_1$ and passing through exactly one vertex in each generation $\mathscr{V}_n$, $n \geq 1$, is called an *end* of $\mathscr{T}$, or an infinite *descent line*. The space of ends of $\mathscr{T}$ (the *limit set* of $\mathscr{G}$) is denoted by Λ; it is endowed with the metric d defined by (1) above. For each vertex v, let $\Lambda(v)$ be the set of all ends that pass through v, and let $\tau(v)$ be the type of v. For each end $\xi = \xi_1\xi_2\cdots$, the sequence of types $\tau(\xi_n)$ will be denoted by $\Phi(\xi)$, and will be called the *pedigree* of the descent line.

Define $\psi : \Omega \to \mathbb{R}$ by

$$\psi(\omega) = \log q(\omega_1, \omega_2) \ .$$

For any Borel probability measure μ on Ω, and any μ-integrable function $f : \Omega \to \mathbb{R}$, we will write

$$E_\mu f = \int_\Omega f \, d\mu \ .$$

Theorem 1. *Let Λ be the limit set of a multi-type Galton–Watson process $\mathscr{G}$ satisfying hypotheses* (*H1*)–(*H2*)*, and let μ be an ergodic, shift-invariant probability measure on Ω with entropy $h(\mu)$. If $h(\mu) + E_\mu\psi < 0$ then the set $\Phi(\Lambda) \cap \Omega_\mu$ is almost surely empty. If $h(\mu) + E_\mu\psi \geq 0$ then almost surely on the event of non-extinction the set $\Lambda \cap \Phi^{-1}(\Omega_\mu)$ has Hausdorff dimension*

$$\delta_H(\Lambda \cap \Phi^{-1}(\Omega_\mu)) = \frac{h(\mu) + E_\mu\psi}{\log 2} \ . \tag{3}$$

The next result concerns the Hausdorff dimensions of sets $\Lambda \cap \Phi^{-1}(U)$, where $U \subset \Omega$ is a set defined by weaker constraints on ergodic averages. Let $\sigma : \Omega \to \Omega$ be the forward shift mapping, let $f : \Omega \to \mathbb{R}^k$ be a continuous vector-valued function, and let J be a nonempty compact subset of $\mathbb{R}^k$. Call a σ-invariant probability measure μ *admissible* if it is ergodic and if $\mu(\Omega(ij)) = 0$ for every pair ij of types for which $q(i, j) = 0$ (note that no such pair can appear in sequence in the pedigree of an infinite descent line).

Define

$$\begin{aligned} S_n f &= f + f \circ \sigma + f \circ \sigma^2 + \cdots + f \circ \sigma^{n-1}, \\ A_n f &= S_n f / n, \\ U(f, J) &= \left\{ \omega \in \Omega \, : \, \lim_{n \to \infty} \text{distance}(A_n f(\omega), J) = 0 \right\}, \qquad \text{and} \\ \mathscr{M}(f, J) &= \left\{ \text{admissible } \mu \, : \, E_\mu f \in J \right\} \ . \end{aligned}$$

Theorem 2. *Let $f : \Omega \to \mathbb{R}^k$ be a function such that, for some finite integer m, $f(\omega)$ is a function only of the first m entries of the argument ω. Let J be a nonempty compact subset of $\mathbb{R}^k$ such that, for every $x \in J$, there exists an admissible probability measure μ on Ω such that $E_\mu f = x$. Then the functional $\mu \to h(\mu) + E_\mu\psi$ attains its maximum on $\mathscr{M}(f, J)$,*

$$\max_{\mu\in\mathscr{M}(f,J)}(h(\mu)+E_\mu\psi)<0 \Longrightarrow \Lambda\cap\Phi^{-1}(U(f,J))=\emptyset \; a.s., \quad \textit{and} \tag{4}$$

$$\delta_H(\Lambda\cap\Phi^{-1}(U(f,J)))=\max_{\mu\in\mathscr{M}(f,J)}\left(\frac{h(\mu)+E_\mu\psi}{\log 2}\right)_+ \tag{5}$$

almost surely on the event of survival.

Remarks

(A) In fact, Theorem 2 holds more generally for functions $f : \Omega \to \mathbb{R}^k$ that are Hölder continuous (with respect to the metric d_Ω). The proof, however, is considerably more technical than the proof in the case where f depends only on finitely many coordinates.

(B) Hawkes' theorem is a special case of Theorem 2: Let $\mathscr{L}$, the space of types, be a singleton; let $f \equiv 1$; and let $J = \{1\}$. Since $\mathscr{L}$ is a singleton, there is only one invariant probability measure on $\Omega = \mathscr{L}^{\mathbb{N}}$, and it has entropy zero. Moreover, $\psi \equiv \log\alpha$, where α is the mean offspring number. Thus, (5) reduces to Hawkes' formula.

(C) Suppose that the space $\mathscr{L}$ of types has cardinality greater than 1. Let μ_* be the unique ergodic, σ-invariant probability measure on Ω that maximizes $h(\mu)+E_\mu\psi$ (that there is a unique maximizer follows, e.g., from Theorem 1.22 of [2]). Now let $f \equiv 1$ and $J = \{1\}$; then $U(f, J) = \Omega$, and $\mathscr{M}(f, J)$ is the set of *all* ergodic, σ-invariant probability measures on Ω. Consequently, Theorems 1 and 2 imply that, if $h(\mu_*) + E_{\mu_*}\psi > 0$, then almost surely on the event of nonextinction,

$$\delta_H(\Lambda)=\delta_H(\Lambda\cap\Phi^{-1}(\Omega_{\mu_*})) \; . \tag{6}$$

Thus, the set of infinite descent lines ζ such that $\Phi(\zeta)$ is generic for the maximizing measure μ_* has Hausdorff dimension as large as the entire set of infinite descent lines.

(D) Now let $f : \Omega \to \mathbb{R}$ be any continuous function, and let J be a closed interval of $\mathbb{R}$ *not* containing $E_{\mu_*}f$. Then by Theorem 2 and the fact that $h(\mu) + E_\mu\psi$ is *uniquely* maximized by μ_*,

$$\delta_H(\Lambda\cap\Phi^{-1}(U(f,J)))<\delta(\Lambda) \; . \tag{7}$$

Thus, the set of infinite descent lines ζ such that $\Phi(\zeta)$ is not nearly generic for μ_* has Hausdorff dimension strictly less than the entire set of infinite descent lines. Together with remark (B), this suggests that in an appropriate sense "most" infinite descent lines are "almost" generic for μ_*.

The remainder of the paper will be devoted to the proofs of Theorems 1 and 2.

2. Proof of Theorem 1: upper bound

In this section we shall prove that

$$\delta_H(\Lambda \cap \Phi^{-1}(\Omega_\mu)) \leq (h(\mu) + E_\mu \psi)/\log 2 \qquad \text{and} \tag{8}$$

$$h(\mu) + E_\mu \psi < 0 \Longrightarrow \Lambda \cap \Phi^{-1}(\Omega_\mu) = \emptyset \tag{9}$$

almost surely. As is often the case in Hausdorff dimension calculations, this is the easy half of the argument: we need only find an efficient sequence of open coverings of the target set $\Lambda \cap \Phi^{-1}(\Omega_\mu)$.

Lemma 1. *Let μ be any ergodic, invariant probability measure on $\Omega = \mathscr{L}^{\mathbb{N}}$. Then for every $\varepsilon > 0$ there exist sets $\Gamma_m = \Gamma_m(\mu) \subset \mathscr{L}^m$ such that*

$$\lim_{m\to\infty} m^{-1} \log |\Gamma_m| = h(\mu) \ , \tag{10}$$

$$\lim_{m\to\infty} \sup_{x\in\Gamma_m} |m^{-1} \sum_{j=0}^{m-1} \psi(\sigma^j x) - \int_\Omega \psi \, d\mu| = 0 \ , \tag{11}$$

and such that for every $x = x_1 x_2 \cdots \in \Omega$,

$$x \in \Omega_\mu \quad \Longrightarrow \quad x_1 x_2 \cdots x_m \in \Gamma_m \text{ infinitely often} \ . \tag{12}$$

Note. In fact, there exist sets Γ_m such that (11) and (10) hold, and for which (12) holds with "infinitely often" replaced by "eventually". This stronger statement will not be needed in the subsequent arguments.

Proof. This is a routine consequence of the Shannon-MacMillan theorem and the ergodic theorem (see, e.g., [11]). These two theorems, together with the ergodicity of μ, imply that there exist sets $\Gamma_m \subset \mathscr{L}^m$ such that

$$\lim_{m\to\infty} \mu(\cup_{x\in\Gamma_m} \Omega(x)) = 1 \ , \tag{13}$$

$$\lim_{m\to\infty} \max_{x\in\Gamma_m} |-m^{-1} \log \mu(\Omega(x)) - h(\mu)| = 0 \ , \tag{14}$$

and such that (11) holds. Observe that, by (14), for each $\varepsilon > 0$, if m is sufficiently large then for every $x \in \Gamma_m$,

$$\exp\{-m(1+\varepsilon)h(\mu)\} \leq \mu(\Omega(x)) \leq \exp\{-m(1-\varepsilon)h(\mu)\} \ ,$$

and so

$$|\Gamma_m| \exp\{-m(1+\varepsilon)h(\mu)\} \leq \mu(\cup_{x\in\Gamma_m} \Omega(x)) \leq |\Gamma_m| \exp\{-m(1-\varepsilon)h(\mu)\} \ ,$$

which, in view of (13), implies (10). To see that (12) must hold, take a subsequence m_k of $\mathbb{N}$ such that

$$\sum_{k=1}^{\infty} (1 - \mu(\cup_{x\in\Gamma_{m_k}} \Omega(x))) < \infty \ .$$

By the Borel-Cantelli Lemma, for μ-almost every sequence $x = x_1 x_2 \cdots \in \Omega$, the prefix $x_1 x_2 \cdots x_{m_k}$ is an element of Γ_{m_k} for all sufficiently large k. □

For any nth generation individual $x \in \mathscr{V}_n$, define the *pedigree* $\Phi(x)$ of x to be the length-n sequence $w = w_1 w_2 \cdots w_n$ recording the types of x and its ancestors; thus, w_k is the type of its ancestor in generation k. For any word $w \in \mathscr{L}^n$, define $\mathscr{V}(w) = \Phi^{-1}(\Omega(w))$ to be the set of all nth generation individuals with pedigree w. Observe that

$$\Phi^{-1}(\Omega(w)) = \bigcup_{x \in \mathscr{V}(w)} \Lambda(x) \quad , \tag{15}$$

and that each set $\Lambda(x)$ in this union has diameter (in the metric $d = d_\Lambda$ defined by (1)) less than 2^{-n}. Observe also that for any word $w = w_1 w_2 \cdots w_n$,

$$E|\mathscr{V}(w)| = q(i_*, w_1) \prod_{j=1}^{n-1} q(w_j, w_{j+1}) = q(i_*, w_1) \exp\left(\sum_{j=1}^{n} \psi(\sigma^{j-1} w)\right) \quad . \tag{16}$$

Fix $\varepsilon > 0$, and let $\Gamma_m = \Gamma_m(\mu)$ be as in the statement of Lemma 1. Then by assertion (12), for each $m = 1, 2, \ldots$,

$$\Phi^{-1}(\Omega_\mu) \subset \bigcup_{n=m}^{\infty} \bigcup_{x \in \Gamma_n} \bigcup_{\xi \in \mathscr{V}(x)} \Lambda(\xi) \tag{17}$$

is a covering of $\Phi^{-1}(\Omega_\mu)$ by open sets of diameter less than 2^{-m}. Denote this covering by $\mathscr{C}_m$. Each $\mathscr{C}_m$ includes infinitely many open sets $\Lambda(\xi)$, but by (16), the expected number of these with diameter 2^{-n} is

$$\sum_{x \in \Gamma_n} q(i_*, x_1) \exp\left(\sum_{j=1}^{n} \psi(\sigma^{j-1} x)\right) \quad , \tag{18}$$

which by assertions (10)–(11) of Lemma 1 is, for any $\varepsilon > 0$, no larger than

$$\exp\left\{n\left(h(\mu) + \int \psi \, d\mu + \varepsilon\right)\right\} \tag{19}$$

provided n is sufficiently large.

Proof of (9). Suppose that $-2\varepsilon = h(\mu) + E_\mu \psi < 0$. Then by (19), for m sufficiently large, the expected number of sets $\mathscr{T}(x)$ included in the covering $\mathscr{C}_m$ is smaller than

$$\sum_{n=m}^{\infty} e^{-n\varepsilon} \quad .$$

Since this converges to zero exponentially fast as $m \to \infty$, the Borel-Cantelli Lemma imples that, with probability one, for all sufficiently large m, the covering $\mathscr{C}_m$ is empty. It therefore follows that with probability one,

$$\Lambda \cap \Phi^{-1}(\Omega_\mu) = \emptyset \quad .$$

Proof of (8). Recall that the Hausdorff dimension of a set F is defined to be $\inf\{\alpha : H_\alpha(F) < \infty\}$, where H_α denotes outer α-dimensional Hausdorff measure. Recall also that $H_\alpha(F)$ is the limit as $\rho \to 0$ of $\inf(\sum_{G\in\mathscr{G}} \text{diameter}(G)^\alpha)$, where the infimum is taken over all coverings $\mathscr{G}$ of F by open sets G of diameter less than ρ. By (17) and (19), for any $\alpha > ((h(\mu) + E_\mu\psi)/\log 2$,

$$E \sum_{\Lambda(\xi)\in\mathscr{C}_m} \text{diameter}(\Lambda(\xi))^\alpha \le \kappa(\alpha) < \infty \ ,$$

for some $\kappa(\alpha) < \infty$ independent of m. Consequently, the outer α-dimensional Hausdorff measure of $\Lambda \cap \Phi^{-1}(\Omega_\mu)$ is almost surely finite, and so the advertised inequality (8) must hold almost surely.

3. Proof of Theorem 1: lower bound

3.1. Basic strategy

In this section, we prove the more difficult inequality

$$\delta_H(\Lambda \cap \Phi^{-1}(\Omega_\mu)) \ge (h(\mu) + E_\mu\psi)/\log 2 \tag{20}$$

almost surely on the event of nonextinction. It suffices to prove this for those ergodic, shift-invariant probability measures μ on Ω such that

$$h(\mu) + E_\mu\psi > 0 \tag{21}$$

because for those μ not satisfying this condition the inequality (20) is trivial. Thus, for the remainder of this section we assume that (21) holds.

Recall that the root of the tree $\mathscr{T}$ is of type i_*. By hypothesis (H2), on the event that $\mathscr{T}$ is infinite, it almost surely has infinitely many vertices of type i_*. For each vertex $x \in \mathscr{T}$ of type i_*, the subtree $\mathscr{T}(x)$ containing those vertices descending from x has the same distribution as the entire tree $\mathscr{T}$, since the root vertex of $\mathscr{T}$ has type i_*. It follows by a routine argument that if for some constant $\gamma > 0$ the Hausdorff dimension of $\Lambda \cap \Phi^{-1}(\Omega_\mu)$ exceeds γ with positive probability, then *almost surely* on the event that $\mathscr{T}$ is infinite, the Hausdorff dimension of $\Lambda \cap \Phi^{-1}(\Omega_\mu)$ exceeds γ; thus, $\delta_H(\Lambda \cap \Phi^{-1}(\Omega_\mu))$ is almost surely constant on the event of nonextinction. Consequently, to establish the inequality (20) it suffices to show that with positive probability there is a subtree $\mathscr{T}^*$ of $\mathscr{T}$ whose set Λ^* of ends satisfies

$$\Phi(\Lambda^*) \subset \Omega_\mu \qquad \text{and} \tag{22}$$

$$\delta_H(\Lambda^*) \ge (h(\mu) + E_\mu\psi)/\log 2 \ . \tag{23}$$

3.2. Background

First we collect some needed results concerning the Hausdorff dimension of nearly regular trees and pruned Galton–Watson trees.

3.2.1. Nearly regular trees

Let T be any infinite rooted tree, with vertex set V partitioned by depth (distance from the root vertex): $V = \cup_{n=0}^{\infty} V_n$. Let N_n and M_n be increasing sequences of positive integers. Say that T is $\{M_n\}$*-regular relative to* $\{N_n\}$ if, for every $n \geq 1$, every vertex x at depth N_n has exactly M_{n+1} descendant vertices at depth N_{n+1}. If T is $\{M_n\}$-regular relative to $\{N_n\}$ for a constant sequence $M_n = D$, say that T is D-regular relative to $\{N_n\}$.

Lemma 2. *Let T be an infinite rooted tree with space of ends ∂T. If there is an increasing sequence of integers N_n satisfying $\lim_{n\to\infty} N_{n+1}/N_n = 1$ such that, for some sequence $\{M_n\}$ of positive integers, T is $\{M_n\}$-regular with respect to $\{N_n\}$, then the Hausdorff dimension of ∂T (relative to the usual metric) is*

$$\delta_H(\partial T) = \liminf_{n\to\infty} \frac{\log |V_{N_n}|}{N_n \log 2} = \liminf_{n\to\infty} \frac{\log \prod_{j=1}^{n} M_n}{N_n \log 2} \quad .$$

Proof. Easy exercise. □

3.2.2. Pruning a Galton–Watson tree

Let $T = T^{\mathscr{Z}}$ be the genealogical tree of an ordinary Galton–Watson process $\mathscr{Z}$ with mean offspring number $\alpha > 1$ and survival probability $\rho > 0$. Let the vertex set V of T be partitioned by depth (generation): $V = \cup_{n=0}^{\infty} V_n$.

Pruning Procedure. Let N_m and M_m be sequences of positive integers. The pruning procedure consists of an infinite sequence of "cuts", each depending on the outcome of the earlier cuts. The set of vertices surviving cut n will be denoted by G_n, and $G_0 = V$. For each $n \geq 0$, the set G_{n+1} is defined to be the set obtained from G_n by removing, for each m, all vertices $x \in G_n$ at depth N_m that have fewer than M_m descendant vertices in G_n at depth N_{m+1}.

Note that if the root vertex survives all the cuts, then T has a subtree that is $\{M_m\}$-regular relative to $\{N_m\}$, and whose vertices are all elements of $\cap_{n\geq 0} G_n$.

Lemma 3. *Suppose that there exist positive probabilities p_m such that for each $m \geq 0$,*

$$\sum_{k=0}^{\infty} P\left\{|V_{N_{m+1}}| = k\right\} P\left\{\sum_{j=1}^{k} \zeta_{m+1,j} \geq M_m\right\} \geq p_m \quad , \tag{24}$$

where $\zeta_{m,j}$ are mutually independent Bernoulli-p_m random variables that are independent of the first N_m generations of the Galton–Watson process $\mathscr{Z}$. Then the probability that the root vertex survives all the cuts of the pruning procedure is at least p_0.

Proof. We will show by induction on n that, for each vertex $x \in V_{N_m}$, the conditional probability that $x \in G_n$, given the first N_m generationsof $\mathscr{Z}$, is at least p_m. Since the sets G_n are decreasing in n, it will follow that $x \in \cap_{n=1}^{\infty} G_n$ with conditional probability at least p_m. This will prove the lemma.

For any vertex $x \in V_{N_m}$, the event that x survives the initial cut has probability 1. For each $x \in \cup_m V_{N_m}$ and each $n \geq 1$, define

$$\begin{aligned} Y_x^n &= 1 \quad \text{if } x \in G_n, \\ Y_x^n &= 0 \quad \text{if } x \notin G_n \ . \end{aligned}$$

Then for each pair m, n the random variables $\left\{ Y_x^n \ : \ x \in V_{N_m} \right\}$ are conditionally independent and identically distributed, given the first N_m generations of $\mathscr{Z}$. Moreover, by the induction hypothesis, for each vertex $x \in V_{N_m}$ the conditional probability that $Y_x^n = 1$ is at least p_m. A vertex $x \in V_{N_m}$ survives cut $n + 1$ if and only if

$$\sum_{y \in V_{N_{m+1}}(x)} Y_y^n \geq M_m \ ,$$

where the sum is over the set $V_{N_{m+1}}(x)$ of vertices $y \in V_{N_{m+1}}$ that are descendants of x. By hypothesis (24), this event has conditional probability at least p_m, given the first N_m generations of $\mathscr{Z}$. □

Verification of the hypothesis (24) in paragraph 3.5 below will be based on the following simple lemma. Recall that $\rho > 0$ is the survival probability of the Galton–Watson process $\mathscr{Z}$.

Lemma 4. *Fix* $1 < \beta < \alpha$ *and* $0 < p \leq 1$. *Suppose that to the vertices* $v \in V$ *are attached random variables* ξ_v *so that* $\{\xi_v \ : \ v \in V_n\}$ *are, for each integer* $n \geq 0$, *conditionally independent Bernoulli-*p, *given* $|V_n|$. *Define events*

$$F(n, p, \beta) = \left\{ \sum_{v \in V_n} \xi_v > \beta^n \right\} \ . \tag{25}$$

Then

$$\lim_{n \to \infty} P(F(n, p, \beta)) = \rho \ . \tag{26}$$

Proof. On the event of survival, $|V_n|/\beta^n \to \infty$ almost surely as $n \to \infty$. Conditional on the event $|V_n| = k$, the random variable $\sum_{v \in V_n} \xi_v$ has the binomial distribution with parameters k and p. Therefore, the lemma follows by an easy application of the Chebyshev inequality. □

Remark. The probabilities $P(F(n, p, \beta))$ are nondecreasing in p.

3.3. μ-Generic subtrees

Recall that $\mathscr{L}$ is the set of types of the multi-type Galton–Watson process $\mathscr{G}$ in Theorems 1–2, $\mathscr{T}$ is the genealogical tree of $\mathscr{G}$, and Λ is the space of ends (equivalently, infinite descent lines) of $\mathscr{T}$. Let $\varepsilon_m > 0$ be a sequence of real numbers decreasing to 0 as $m \to \infty$, and let $n_m = n(m) \geq m$ be an increasing sequence of positive integers. Define F_m to be the set of all words $w \in \mathscr{L}^{n(m)}$ such that

(a) w ends with the letter i_*;
(b) w begins with a letter w_1 for which $q(i_*, w_1) > 0$;
(c) no two-letter word $u = ij$ for which $q(i, j) = 0$ appears in w; and
(d) for each word $u \in \cup_{j=1}^{m} \mathscr{L}^j$, the relative frequency of u in w differs from $\mu(\Omega(u))$ by less than ε_m.

Observe that if $u = ij$ is a two-letter word for which $q(i, j) = 0$ then $\mu(\Omega(u)) = 0$, because $E_\mu \psi > -\infty$. Now let $k_m = k(m)$ be any sequence of positive integers. Define $\Lambda' \subset \Lambda$ to be the subset including those ends $\xi \in \Lambda$ for which the pedigree $\Phi(\xi)$ consists of k_1 words in F_1, followed by k_2 words of F_2, etc. The following lemma is an obvious consequence of the assumption that $\lim_{m\to\infty} \varepsilon_m = 0$, together with the definitions of the sets F_m and Ω_μ.

Lemma 5. $\Phi(\Lambda') \subset \Omega_\mu$.

The remainder of the argument entails showing that the as yet unspecified sequences k_m and n_m can be chosen so that with positive probability, $\mathscr{T}$ has a subtree $\mathscr{T}^*$ whose set Λ^* of ends is contained in Λ' and satisfies (23). For this, the following lemma is crucial.

Lemma 6. *For any sequence $\varepsilon_m > 0$ such that $\lim_{m\to\infty} \varepsilon_m = 0$, if $n_m \to \infty$ sufficiently rapidly then*

$$\liminf_{m\to\infty} \frac{1}{n_m} \log \left(\sum_{w \in F_m} \prod_{i=1}^{n(m)-1} q(w_i, w_{i+1}) \right) \geq h(\mu) + E_\mu \psi > 0 \ . \tag{27}$$

Proof. By definition of F_m, for every $w \in F_m$ and every two-letter word u, the relative frequency of u in w differs from $\mu(\Omega(u))$ by less than ε_m. Consequently, there exist constants $\delta_m > 0$, depending only on ε_m and the mean offspring numbers q_{ij}, such that $\lim_{m\to\infty} \delta_m = 0$ and for every $w \in F_m$,

$$\prod_{i=1}^{n(m)-1} q(w_i, w_{i+1}) \geq \exp\left\{ n_m (E_\mu \psi - \delta_m) \right\} \ .$$

The Shannon-MacMillan(-Breiman) Theorem (see, e.g., [11]) and the Ergodic Theorem imply that for any $\varepsilon_m > 0$, if n_m is sufficiently large then there is a subset F_m^* of F_m such that

$$\sum_{w \in F_m^*} \mu(\Omega(w)) \geq \mu(\Omega(i_*))/2$$

and such that for all $w \in F_m^*$,

$$\mu(\Omega(w)) \leq \exp\{-n_m(h(\mu) + \delta_m)\} \quad .$$

It follows that

$$|F_m| \geq \exp\{n_m(h(\mu) - \delta_m)\}\, \mu(\Omega(i_*)/2 \quad .$$

This, together with the inequality of the preceding paragraph, implies the inequality (27), since $\lim_{m\to\infty} \delta_m = 0$. □

Remark. The sequence $\{n_m\}_{m\geq 1}$ may always be chosen so as to be strictly increasing. Moreover, it may be replaced by the sequence $\{n_{m+k}\}_{m\geq 1}$, and so it may be arranged that the infimum of the sequence on the left side of (27) may be made arbitrarily close to $h(\mu) + E_\mu \psi$. Since by hypothesis this is positive, we may choose $\{n_m\}_{m\geq 1}$ so that, for every $m \geq 1$,

$$\alpha_m \triangleq \sum_{w\in F_m} \prod_{i=0}^{n(m)-1} q(w_i, w_{i+1}) > 1 \quad , \tag{28}$$

where $w_0 = i_*$.

3.4. Embedded Galton–Watson processes

Recall that the vertex set $\mathscr{V}$ of $\mathscr{T}$ is partitioned by depth: $\mathscr{V} = \cup_{n=0}^{\infty} \mathscr{V}_n$; and that $\mathscr{V}_0$ consists of a single vertex of type i_*. For each $m \geq 1$ define subsets $U_j^m \subset \mathscr{V}_{jn(m)}$ inductively as follows:

(a) $U_0^m = \mathscr{V}_0$; and
(b) for each vertex $x \in \mathscr{V}_{(j+1)n(m)}$, x is included in U_{j+1}^m if and only if
  (i) it is a descendant (in $\mathscr{T}$) of a vertex in U_j^m; and
  (ii) the last n_m letters of the pedigree $\Phi(x)$ form a word in F_m.

Lemma 7. *For each m, the sequence of random variables $(|U_j^m|)_{j\geq 0}$ is an* ordinary *Galton–Watson process with mean offspring number α_m defined by* (28).

Proof. Every vertex $x \in U_j^m$ has a pedigree $\Phi(x)$ ending in the letter i_*. Consequently, for each such x, the set $\mathscr{T}(x)$ of descendants of x in $\mathscr{T}$ constitutes a realization of the multitype Galton–Watson process $\mathscr{G}$. Moreover, these realizations are conditionally independent, given the first jn_m generations of $\mathscr{G}$, since $\mathscr{G}$ is a (multi-type) Galton–Watson process. Thus, $(|U_j^m|)_{j\geq 0}$ is an ordinary Galton–Watson process. That the mean offspring number is given by the sum in (28) is apparent from the construction. □

For each m, the mean offspring number α_m of the Galton–Watson process $|U_j^m|$ is, by (28), larger than 1; hence, $|U_j^m|$ is supercritical. Define

$$\rho_m = \rho(m) = P\left\{\lim_{j\to\infty} |U_j^m| = \infty\right\} \tag{29}$$

to be its survival probability.

3.5. Existence of nearly regular subtrees

Using the embedded Galton–Watson processes defined in paragraph 3.4, we will show that integers k_m may be chosen so that if Λ' is defined as in paragraph 3.3, then with positive probability $\mathscr{T}$ has a subtree $\mathscr{T}^*$ whose set Λ^* of ends is contained in Λ' and satisfies (23). This will complete the proof of (20) and, hence, that of Theorem 1.

Define

$$p_m = p(m) = \min(\rho_m, \rho_{m+1})/2 \ . \tag{30}$$

By Lemma 6, there exist integers β_m so that for each $m \geq 1$,

$$\beta_m < \alpha_m \qquad \text{and} \tag{31}$$

$$\lim_{m\to\infty} \frac{\log \beta_m}{n_m} = h(\mu) + E_\mu \psi \ . \tag{32}$$

Now suppose, as in Lemma 4, that to each vertex $v \in \cup_j U_j^m$ is attached a random variable Y_v, so that for each generation j the random variables Y_v, where $v \in U_j^m$, are conditionally independent given $|U_j^m|$, each with the Bernoulli-p_m distribution. By Lemma 4, there exist integers $r_m = r(m)$ such that for each m,

$$P(F_m(r_m, p_m, \beta_m)) > \rho_m/2 \tag{33}$$

where

$$F_m(r, p_m, \beta_m) = \left\{ \sum_{v \in U_j^m} Y_v > \beta_m^r \right\} \ .$$

We may assume that the sequence r_m is increasing. Finally, let $k_m = s_m r_m$ be an integer multiple of r_m, with s_m chosen so that

$$\lim_{m\to\infty} \frac{r_{m+1} n_{m+1}}{k_m n_m} = 0 \ . \tag{34}$$

Lemma 8. *If the sequences r_m, s_m, and $k_m = s_m r_m$ satisfy* (33) *and* (34)*, and if Λ' is defined as in paragraph* 3.3*, then with positive probability $\mathscr{T}$ has a subtree $\mathscr{T}^*$ whose set Λ^* of ends is contained in Λ' and has Hausdorff dimension at least $h(\mu) + E_\mu \psi$.*

Proof. For each positive integer j let $m = m(j)$ be the unique positive integer such that

$$\sum_{i=1}^{m-1} s_i < j \leq \sum_{i=1}^{m} s_i \ ,$$

and define

$$M_j = \beta_m^{r_m} \qquad \text{and}$$

$$N_j = \sum_{i=1}^{m-1} k_i n_i + \left(j - \sum_{i=1}^{m-1} s_i \right) r_m n_m \ .$$

Relations (34) and (32) imply that

$$\lim_{j\to\infty} \frac{N_{j+1}}{N_j} = 1 \qquad \text{and} \tag{35}$$

$$\lim_{j\to\infty} \frac{\log \prod_{i=1}^{j} M_i}{N_j \log 2} = \frac{h(\mu) + E_\mu \psi}{\log 2} \quad . \tag{36}$$

Therefore, by Lemma 2, it suffices to prove that with positive probability $\mathscr{T}$ contains a subtree $\mathscr{T}^*$ whose set of ends is contained in Λ' and which is $\{M_j\}$-regular with respect to $\{N_j\}$.

Let x be a vertex of $\mathscr{T}$ at depth N_j. Partition the descendants y of x at depth N_{j+1} into two classes, "good" and "bad": say that y is *good* if the last $N_{j+1} - N_j$ letters of its pedigree are the concatenation of r_m words in F_m (if $N_{j+1} - N_j = r_m n_m$) or the concatenation of r_{m+1} words in F_{m+1} (if $N_{j+1} - N_j = r_{m+1} n_{m+1}$); otherwise, say that y is *bad*. Define Y_x to be 1 if x has at least M_m good descendants at depth N_{j+1}, and 0 otherwise. Then by (33), the conditional probability that $Y_x = 1$, given the first N_j generations of the Galton–Watson process $\mathscr{G}$, is at least p_m. Consequently, by (33) and Lemma 3, there is positive probability that the root vertex survives all cuts of the pruning procedure described in paragraph 3.2.2. On this event, the pruned tree contains a subtree that is $\{M_j\}$-regular with respect to $\{N_j\}$. By construction, the limit set of the pruned tree is contained in Λ'. $\square$

4. Proof of Theorem 2

Observe first that, by Theorem 1, $\delta_H(\Lambda \cap \Phi^{-1}(U(f, J))$ is, on the event of survival, at least as large as the maximum over $\mathscr{M}(f, J)$ of $(h(\mu) + E_\mu \psi)/\log 2$, because, for each $\mu \in \mathscr{M}(f, J)$, the set Ω_μ is contained in $U(f, J)$. Thus, to prove equation (5), it suffices to prove the inequality

$$\delta_H(\Lambda \cap \Phi^{-1}(U(f, J))) \leq \max_{\mu \in \mathscr{M}(f,J)} \left(\frac{h(\mu) + E_\mu \psi}{\log 2} \right)_+ \quad . \tag{37}$$

Without loss of generality, we may assume that the function $f : \Omega \to \mathbb{R}^k$ is a function only of the first two entries of its argument. This is because, for any integer $k \geq 1$, the type space $\mathscr{L}$ of the underlying multi-type Galton–Watson process $\mathscr{G}$ may be enlarged to $\mathscr{L}^m$, with each individual of $\mathscr{G}$ being re-labelled in accordance with its type and the types of its $m - 1$ most recent ancestors. Moreover, we may assume that the components of the vector f are the indicator functions f_{ij} of those pairs ij of types for which $q(i, j) > 0$. (If f is not originally of this form, it can be transformed by adjoining indicator functions, replacing components by linear combinations of components, and deleting redundant components. None of these transformations affects the truth of the statements in Theorem 2.)

Say that a σ-invariant probability measure μ on Ω is *Markov* if, under μ, the coordinate process is a (1-step) Markov chain, equivalently, if for every finite sequence $x_1 x_2 \cdots x_m$ such that $\mu(\Omega(x_1 x_2 \cdots x_m) > 0$,

$$\frac{\mu(\Omega(x_1 x_2 \cdots x_m))}{\mu(\Omega(x_1 x_2 \cdots x_{m-1}))} = \frac{\mu(\Omega(x_{m-1} x_m))}{\mu(\Omega(x_{m-1}))} \quad . \tag{38}$$

Lemma 9. *If, for some $w \in \mathbb{R}^k$, there exists an admissible, σ-invariant probability measure ν on Ω such that $E_\nu f = w$, then there exists an admissible Markov measure $\mu = \mu_w$ such that $E_\mu f = w$, and*

$$h(\mu) \geq h(\nu) \quad . \tag{39}$$

Proof. For every pair $ij \in \mathscr{L}^2$ define $\pi(ij) = \nu(\Omega(ij))$ and $\pi(i) = \nu(\Omega(i))$. Then there is a unique Markov measure μ on Ω with the same marginals $\pi(ij)$ (the transition probabilities are $p(i, j) = \pi(ij)/\pi(i)$ for those pairs ij such that $\pi(ij) > 0$, and 0 for the rest). Since the components of f are just the indicators of the different allowable pairs ij of types, $E_\mu f = E_\nu f$. Since ν is admissible, so is μ.

The inequality (39) is a routine consequence of the Jensen inequality for convex functions. Letting $\varphi(t) = -t \log t$, we have

$$\begin{aligned} h(\nu) &= \lim_{n\to\infty} \frac{1}{2} \sum_{x_1 x_2 \cdots x_n} \mu(\Omega(x_1 x_2 \cdots x_{n-2}))\varphi\left(\frac{\mu(\Omega(x_1 x_2 \cdots x_n))}{\mu(\Omega(x_1 x_2 \cdots x_{n-2}))}\right) \\ &\leq \lim_{n\to\infty} \frac{1}{2} \sum_{x_{n-1} x_n} \varphi\left(\sum_{x_1 x_2 \cdots x_{n-2}} \mu(\Omega(x_1 x_2 \cdots x_n))\right) \\ &= \lim_{n\to\infty} \frac{1}{2} \sum_{x_{n-1} x_n} \varphi(\mu(\Omega(x_{n-1} x_n))) \\ &= h(\mu) \quad . \end{aligned}$$

□

One of the advantages of dealing with Markov measures is that the entropy functional $h(\mu)$ varies continuously with the transition probabilities p_{ij}. Hence, it follows from Lemma 9 that, for any nonempty compact set $J \subset \{E_\mu f \;:\; \mu \text{ admissible}\}$, the functional $\mu \to E_\mu f + h(\mu)$ attains its maximum in $\mathscr{M}(f, J)$, and does so at an admissible Markov measure.

Lemma 10. *For each admissible Markov measure $\mu = \mu_w$ and each $\varepsilon > 0$ there exists $\delta > 0$ such that for all sufficiently large n,*

$$|\left\{x \in \mathscr{L}^n \;:\; |A_n f(x) - w| \leq \delta\right\}| \leq \exp\{nh(\mu) + n\varepsilon\} \quad . \tag{40}$$

Proof. Consider the set $\mathscr{N}(\mu) = \mathscr{N}_\delta(\mu)$ of all Markov measures μ' for which the two-step marginal probabilities $\mu'(\Omega(ij))$ are all within δ of the corresponding marginal probabilities $\mu(\Omega(ij))$ for μ. For any $\mu' \in \mathscr{N}(\mu)$, the marginal probabilities $\mu'(\Omega(ij))$ are bounded above by $\nu(i, j)$, where

$$\nu(i, j) = \mu(\Omega(ij)) + \delta \quad .$$

Define transition probabilities $p_*(i, j)$ by

$$\begin{aligned} p_*(i, j) &= \nu(i, j)/\nu(i), \\ \nu(i) &= \sum_j \nu(i, j) \quad . \end{aligned}$$

Let μ_* be the Markov measure for which the associated Markov chain has transition probabilities $p_*(i, j)$. Observe that, since $\nu(i, j) > 0$, the Markov measure μ_* is mixing. If $\delta > 0$ is small then the transition probabilities $p_*(i, j)$, and consequently also the stationary probabilities $\pi_*(i)$, are close to those of the Markov chain asociated with the Markov measure μ. It follows that if $\delta > 0$ is small then the differences $|\pi_*(i) - \nu(i)|$ and $|h(\mu_*) - h(\mu)|$ are small.

For any sequence $x \in \mathscr{L}^n$ such that $|A_n f(x) - w| < \delta$, the empirical distribution of adjacent pairs ij in x must coincide with the two-step marginal probabilities $\mu'(\Omega(ij))$ for some $\mu' \in \mathscr{N}(\mu)$. Consequently, by the definitions of the quantities ν and $p_*(i, j)$, for each such sequence $x = x_1 x_2 \cdots x_n$,

$$\begin{aligned}\prod_{m=1}^{n-1} p_*(x_m, x_{m+1}) &\geq \prod_{i,j} p_*(i, j)^{n\nu(i,j)} \\ &\geq e^{-n\varepsilon/4} \prod_{i,j} p_*(i, j)^{n\pi_*(i)p_*(i,j)} \\ &\geq \exp\{-nh(\mu_*) - n\varepsilon/3\} \\ &\geq \exp\{-nh(\mu) - n\varepsilon/2\} \quad ,\end{aligned}$$

provided $\delta > 0$ is sufficiently small and n is large. Since

$$\sum_{x \in \mathscr{L}^n} \pi_*(x_1) \prod_{m=1}^{n-1} p_*(x_m, x_{m+1}) = 1 \quad ,$$

it follows that the number of sequences $x \in \mathscr{L}^n$ such that $|A_n f(x) - w| < \delta$ cannot exceed $\exp\{nh(\mu) + n\varepsilon\}$ for large n. □

Proof of Inequalities (4) and (37). Let $J \subset \mathbb{R}^k$ be a compact set such that for every $w \in J$ there exists an admissible probability measure μ for which $E_\mu f = w$. By Lemma 9, the measure μ_w that maximizes entropy subject to the constraint $E_\mu f = w$ is a Markov measure. By Lemma 10, for each $\varepsilon > 0$ and each $w \in J$ there exists $\delta = \delta_w > 0$ such that for all large n, inequality (40) holds, and so that, for each $x = x_1 x_2 \cdots x_n \in \mathscr{L}^n$ such that $|A_n f(x) - w| < \delta_i$,

$$q(i_*, x_1) \prod_{j=1}^{n-1} q(x_j, x_{j+1}) \leq \exp\left\{n \int \psi \, d\mu_w + n\varepsilon\right\} \quad . \tag{41}$$

Since J is compact, there exists a finite subset $\{w_1, w_2, \ldots, w_r\}$ such that J is covered by the balls B_i of radii δ_{w_i} centered at the points w_i. Hence,

$$U(f, J) \subset \cup_{i=1}^r U(f, B_i) \quad .$$

As in section 2, for any $x \in \mathscr{L}^n$ let $\mathscr{V}(x) = \Phi^{-1}(\Omega(x))$ be the set of all nth generation vertices of $\mathscr{T}$ with pedigree x. Then for each $m \geq 1$,

$$U(f, B_i) \subset \bigcup_{n=m}^{\infty} \bigcup_{x \in \Delta(i,n)} \bigcup_{\xi \in \mathscr{V}(x)} \Lambda(\xi)$$

where

$$\Delta(i,n) = \left\{ x \in \mathscr{L}^n \ : \ |A_n f(x) - w_i| < \delta_{w_i} \right\} \quad .$$

This is a covering of $U(f, B_i)$ by sets of diameter no larger than 2^{-m}. By (16), (41), and (40), the expected number of sets of diameter 2^{-n} in this covering is no larger than

$$\exp\left\{ n(h(\mu_{w_i}) + \int \psi \, d\mu_{w_i} + 2\varepsilon) \right\} \quad .$$

The proofs of (4) and (37) may now be completed by the same arguments as in section 2.

References

1. Athreya, K., and Ney, P.: Branching Processes. Springer-Verlag (1972)
2. Bowen, R.: Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms. Springer Lecture Notes in Math. **470**, (1975)
3. Evans, S.: Polar and nonpolar sets for a tree indexed process. Annals of Probability **20**(2), 579–590 (1992)
4. Furstenberg, H.: Intersections of Cantor sets and transversality of semigroups. Problems in analysis (Bochner Symp.) 41–59 (1970)
5. Hawkes, J.: Trees generated by a simple branching process. J. London Math. Soc.(2) **24**, 373–384 (1981)
6. Lalley, S., Hueter, I.: Anisotropic branching random walks on homogeneous trees. To appear, Probability & Related Fields (MS1247) (1998)
7. Lalley, S., Sellke, T.: Limit set of a weakly supercritical contact process on a homogeneous tree. Annals of Probability **26**, 644–657 (1998)
8. Lyons, R.: Random walks and percolation on trees. Annals of Probability **18**, 931–958 (1990)
9. Pemantle, R., Peres, Y.: Galton-Watson trees with the same mean have the same polar sets. Annals of Probability **23**(3), 1102–1124 (1995)
10. Pesin, Y.: Dimension Theory in Dynamical Systems: Contemporary Views and Applications Universtiy of Chicago Press, Chicago IL (1997)
11. Petersen, K.: Ergodic Theory. Cambridge Univ. Press, Cambridge UK (1983)
12. Walters, P.: An Introduction to Ergodic Theory. Springer-Verlag, NY (1982)