Chapter 4

EDGE-REINFORCED RANDOM WALK

ABSTRACT

A random walk on a graph is given a particular kind of positive feedback so edges already traversed are more likely to be traversed in the future. Using exchangeability theory, the process is shown to be equivalent to a random walk in a random environment, that is to say, a mixture of Markov chains. When the graph is finite, the fractional occupation times for the edges approach random limits whose joint distributions can be calculated explicitly. When the graph is infinite and acyclic, the process can vary from transient to positive recurrent, depending on the parameter measuring the strength of the reinforcement and on the rate of growth of the graph. The values of the parameters at this phase transition are calculated.

4.1 Introduction

The idea of edge-reinforced random walk is due to Coppersmith and Diaconis. Imagine a person getting acquainted with a new town. She walks about the area near the hotel somewhat randomly, but tends to traverse the same block over and over as they become familiar. To model this, Coppersmith and Diaconis (1987) have defined the following process which they call Reinforced Random Walk. A random walk is taken on the vertices of an undirected graph, beginning at a specified vertex. Initially all the edges are given weight 1, but whenever an edge is traversed the weight of that edge is increased by a fixed parameter, Δ. To choose the next move from a particular vertex, an edge leading out from the vertex is chosen, with the probabilities for the various edges being proportional to their weights. So for example, if after one step the walk has reached a vertex with k neighbors, it will return to the starting point on the next step with probability $(1 + \Delta/(k + \Delta)$. For a more formal description of edge-reinforced random walk, see [CD].

In section 2, we present an analysis of this process on a general finite graph that is due to [CD]. They use the notion of partial exchangeability to equate the process with a mixture of Markov chains and calculate the random limiting fraction of the time that the walk spends on each edge. The rest of this chapter studies the case where the graph is an infinite tree (acyclic graph). The starting point of the investigation is a mapping of the edge-reinforced random walk into a random walk in a random environment determined by infinitely many independent Pólya urns. We apply this to the sequence of edges chosen each time the walk is at a fixed vertex. Since the infinite graphs considered are acyclic, the full force of partial exchangeability is not needed; we use only a single result from the introduction, namely theorem 2.2.

In section 4 Pólya's Urn is used to construct a random walk in a random environment (RWRE) that is equivalent to the original edge-reinforced random walk. In sections 5 and 6 we study RWRE using a large deviation bound from Chernoff's original paper [Ch]. Section 5 contains a sufficient criterion for a.s. transience of a certain class of RWRE. Section 6 gives a sufficient condition (via the stationary measure) for positive recurrence of the RWRE. The only cases remaining unsettled are transitional points where certain equalities hold. We apply these results in section 7 to the RWRE described in section 4 preceding lemma 4.7. surprisingly, the calculations for this class of RWRE reduce to a few lines. Thus the recurrence or transience of edge-reinforced random walk on an infinite binary tree can be established except for one value of Δ.

4.2 Finite graphs

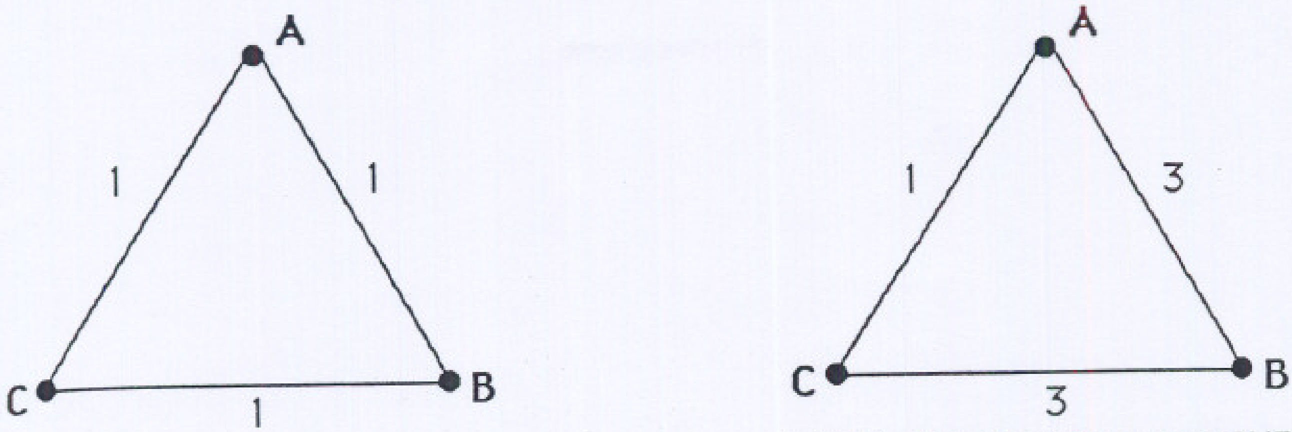
The results of Coppersmith and Diaconis are valid for any finite graph, but for brevity and concreteness we will consider the case where the graph is a triangle. We will also assume that $\Delta = 1$. The reader is referred to [CD] or [Di] for details about the general case. Let the vertices of the triangle be labelled A, B and C and let the edges be labelled BC, CA and AB (see figure 1). The sequence of vertices visited by the reinforced random walk is certainly not exchangeable. It is possible, for example, that the sequence begins ABACBCBA, but it is impossible to begin with an arbitrary permutation of these, say AAABBBCC, because no vertex can be visited twice in a row. However, there holds a certain partial exchangeability, where a partially exchangeable sequence is precisely one for which the next proposition holds.

Proposition 4.1 ([CD]) *Let $V_1, \ldots, V_n$ and $W_1, \ldots, W_n$ be two sequences of vertices of the triangle such that they have the same starting points*

$$V_1 = W_1$$

and the same transition counts

$$card\{k : V_k = i, V_{k+1} = j\} = card\{k : W_k = i, W_{k+1} = j\}$$



INITIALLY THE WALK IS AT VERTEX A

AFTER THE SEQUENCE OF MOVES
A B A C B C B A

FIGURE 1

for all vertices i and j. Then the edge-reinforced random walk starting at V_1 has the same probability of beginning with the sequence $V_1, \ldots, V_n$ as with the sequence $W_1, \ldots, W_n$.

Proof: The probability that the sequence of vertices visited begins $V_1, \ldots, V_n$ is a product of $n-1$ terms whose numerators are the number of times an edge has been traversed so far and whose denominators are the number of times all edges adjacent to a certain vertex have been traversed so far. The product of the numerators may be calculated from the transition counts alone, since an edge that gets traversed k times contributes a factor of $k!$ to the numerator. Similarly, the denominator is determined by the transition counts, so the probability in question is determined solely by the transition counts. □

We can then apply a theorem or Diaconis and Freedman characterizing partially exchangeable processes as mixtures of Markov chains. For full generality on partial exchangeability with a countable state space, see [DF].

Theorem 4.2 ([DF]) *Let $X_1, X_2, \ldots$ be a recurrent partially exchangeable sequence of random variables with values in a finite domain. Then the sequence is a mixture of Markov chains. More precisely, there is a measure μ on transition matrices for the finite domain, such that*

$$prob(X_1 = V_1, \ldots, X_n = V_n) = \int \left(\prod_{j=1}^{n-1} M_{V_j, V_{j+1}} \right) d\mu(M).$$

□

As usual, the empirical transition matrices (with the rows normalized to sum to 1) converge almost surely to a matrix M with distribution μ so the sample value of M can in some sense be recovered in the limit. Combining proposition 4.1 with theorem 4.2

and an easy verification that the vertex-reinforced random walk is recurrent, we see that it is a mixture of Markov chains. To describe the mixing measure, it is easier to describe the joint distribution of the fractions of time the walk spends on each edge. It is clear form the mechanism of updating the transition probabilities that the sample value of M carries the same information as the limiting occupation fraction for all the edges.

If the limiting fractional occupation of each edge is made into a vector, then the vector must lie on the unit simplex because the coordinates must sum to 1. Coppersmith and Diaconis give the density of this vector with respect to the area measure on the simplex. Let x, y and z be the limiting fractional occupations of edges BC, CA and AB respectively. Assume the walk begins at vertex A. Their formula simplifies to

$$prob((x,y,z) \in \mathcal{S}) = \int_{\mathcal{S}\cap\Delta} (xy+yz+zx)^{1/2}(y+z)^{-1}(z+x)^{-3/2}(x+y)^{-3/2}dA.$$

The general formula is similar but it includes a polynomial that is a determinant of a matrix indexed by the homology group of the graph. Their result for finite trees is a little simpler because the homology vanishes. To describe the density of the limiting fractional occupation vector in this case, let each edge e have an associated variable x_e. For any vertex v let x_v denote the sum of x_e over all edges e that are incident to v. Let r_e be the initial edge weights, let s_v be the sum of r_e over all e incident to v and let v_0 be the starting vertex. Then the density with respect to the area measure on the unit simplex is

$$C \prod_e {x_e}^{r_e - 1/2} \prod_v {x_v}^{-(s_v+1)/2} {x_{v_0}}^{1/2} \tag{1}$$

for some constant C. From this it is possible to calculate the mixing measure μ over transition matrices in the representation of this process as a mixture of Markov chains. In the next section it will be shown that the rows of the transition matrix are independent and Dirichlet distributed, but this is not at all apparent from equation (1). The derivation in [CD] of (1) involves a lot of calculus and is too complicated for me to write down or you to want to read.

4.3 Main results for infinite trees

Say the edge-reinforced random walk is recurrent iff the probability of return to the root is 1. After proposition 4.8 of the next section, it will be clear that the usual equivalences hold: the walk is recurrent iff it returns to the root infinitely often a.s. , and it is transient iff it returns to the root finitely often a.s.

Our notation for trees is as follows. The set of vertices or nodes is a finite or countable set, $\mathbf{T}$. The starting node, or root, is denoted ρ. Every node v other than ρ is adjacent to a parent, $\mathrm{par}(v)$, which is closer to the root, and zero or more children, denoted $c_1(v), c_2(v)$, etc. We will write $v_1 \leq v_2$ for v_1 an ancestor of v_2. A branch of length $n \leq \infty$ is a sequence of nodes of length n beginning with ρ where each is the parent of the next. $\mathbf{T}_n$ denotes the set of nodes at distance n from ρ.

The mean recurrence time is always infinite. To see this for $\Delta \geq 1$, let v_1 be one of k vertices adjacent to the root and let v_2 be adjacent to v_1 and distinct from the root. Then the probability of going from v_1 to v_2 at least $n+1$ times before returning to the root is at least

$$\begin{aligned} &(1/k)\ \ (1/(2+\Delta))\ \ ((1+2\Delta)/(2+3\Delta))\ \cdots \\ &\times((1+2n\Delta)/(2+(2n-1)\Delta)) \\ &\geq 1/k(2+(2n-1)\Delta)\,. \end{aligned} \tag{2}$$

The mean recurrence time is

$$\begin{aligned} &\sum_{n=1}^{\infty} n\,prob(\text{first return at time } n) \\ =\ &\sum_{n=0}^{\infty} prob(\text{not returned by step } n) \\ \geq\ &\sum_{n=0}^{\infty} \frac{1}{k(2+(2n-1)\Delta)} \end{aligned}$$

which diverges. Note that this result holds even if the tree itself is finite. If $\Delta < 1$ the mean recurrence time is also infinite but we need lemma 4.7 below to see this.

Nevertheless, we can define a notion of mixed positive recurrence. In section 4 the walk is decomposed into a mixture of Markov chains, the mixture being necessarily unique in the recurrent case. Call the walk "mixed positive recurrent" whenever the Markov chain is a.s. positive recurrent under the mixing measure.

Theorem 4.3 *For an edge-reinforced random walk on an infinite binary tree, there exists*
$\Delta_0 \approx 4.29$ *so that*

$$\text{For } \Delta < \Delta_0 \text{ the walk is transient;} \tag{3}$$

$$\text{For } \Delta > \Delta_0 \text{ the walk is mixed positive recurrent.} \tag{4}$$

It should be noted that the author does not know whether increasing Δ always makes an edge-reinforced random walk more recurrent in any quantitative sense. It seems reasonable to conjecture that the probability of return to the root is monotone in Δ.

In more generality, we can allow the tree itself to be random as long as everything is sufficiently i.i.d. .

Definition 4.4 *By* an i.i.d. RWRE on a random tree *we mean the following. Let each vertex v have $M(v)$ children with $M(v)$ i.i.d. and bounded and $\mathbf{E}(M) = \lambda > 1$. Writing a_i for prob$(M(v) = i)$, this means that $\sum a_i = 1$ and $\sum i a_i = \lambda$. For binary trees $M(v) \equiv \lambda = 2$. In general, the tree can be any supercritical Galton-Watson process. Let the transition probabilities from v to its neighbors $v_1, v_2, \ldots, v_M$ be denoted by the vector $\vec{p}$ with p_0 being the probability of transition to the parent of v and the conditional distribution of $\vec{p}$ given M being symmetric in coordinates $p_1, \ldots, p_M$. The collection of random variables $\{\vec{p}(v) : v \in \mathbf{T}\}$ should be independent after conditioning on the* *shape of the tree, and should be identically distributed for nodes with the same number of children.*

Let

$$\phi(v) = \textit{prob}\,(\text{transition from parent of } v \text{ to } v) \div \tag{5}$$

False

$$\textit{prob}\,(\text{transition from parent of } v \text{ to grandparent of } v).$$

So ϕ is the same on the class of all children of the same node and i.i.d. on such equivalence classes except at the root and any of its children. (As a harmless fiction we sometimes pretend these exceptions do not exist.)

References will be made to expectations involving ϕ and the following is a more precise statement of the measure against which these expectations are to be taken. Let μ_i for $i = 1, 2, 3, \ldots$ be the law of $\phi(v)$ conditioned on $\text{par}(v)$ having exactly i children, i.e. the law of $p_1(\text{par}(v))/p_0(\text{par}(v)) given M(\text{par}(v)) = i$. This measure is well defined because the distributions of $\vec{p}(v)$ are supposed to be identical for nodes with the same number of children. Then the measure we take to be the "law of ϕ" is

$$\mu = \sum_i i\, a_i\, \mu_i / \lambda.$$

This formula corresponds to picking a vertex uniformly from some generation and looking at ϕ for that vertex. The distribution of ϕ will depend on the number of children of the parent of the chosen vertex. The probability of the parent having i children is not a_i but ia_i/λ because nodes with more children are more likely to get picked as parents of a node chosen uniformly.

Let

$$m(r) = \inf\{\exp(-rt)\mathrm{E}(\phi^t) : t \in \mathbf{R}\} \tag{6}$$

be the rate function for $\ln(\phi)$ as in (22) below. Assume that $\mathrm{E}(\ln(\phi))$ exists, possibly $\pm\infty$. The following theorem collects all results on i.i.d. RWRE.

Theorem 4.5 *Conditional on the tree being infinite:*

If	$\mathbf{E}(\ln(\phi)) \geq 0$ *then the walk is a.s. transient;*	(7)
If	$\mathbf{E}(\ln(\phi)) < 0$ *and* $\sup\{\lambda r \, m(\ln(r)) : 0 < r \leq 1\} < 1$ *then the walk is a.s. positive recurrent;*	(8)
If	$\mathbf{E}(\ln(\phi)) < 0$ *and* $\sup\{\lambda r \, m(\ln(r)) : 0 < r \leq 1\} > 1$ *then the walk is a.s. transient;*	(9)
If	$\mathbf{E}(\phi) < 1/\lambda$ *then the walk is a.s. positive recurrent.*	(10)
If	$1 \leq \mathbf{E}(\phi) \leq \infty$ *then the suprema in (8) and (9) need only be evaluated at* $r = 1$.	(11)

Boundedness of M is not really needed except in (8); even here it may be replaced by a weaker condition. (10) is always included in (8) but is given for ease of calculation. (7) is included because the calculation is easier than (9) but the case $\mathbf{E}(\ln(\phi))$ can always be removed by adding "ghost" children to each vertex that have zero probability of ever being reached. Here are two examples to illustrate the various parts of the theorem.

Example: Let $M(v)$ have any distribution that makes the tree a supercritical Galton-Watson process and let γ be a positive real parameter. Suppose that a vertex with k children always has $p_0 = \gamma/\gamma + k$ and $p_i = 1/\gamma + k$ for $i \geq 1$. Then $\phi \equiv 1/\gamma$. If $\gamma \leq 1$ then $\mathbf{E}(\ln(\phi)) \geq 0$ and the walk is transient by (7). Otherwise, use the fact that

$$m(r) = \begin{cases} 1 \text{ if } r = \ln(1/\gamma) \\ 0 \text{ otherwise} \end{cases} .$$

Then $\sup \lambda r m(\ln(r)) = \lambda/\gamma$ so λ is a critcal value for γ with transience when $\gamma < \lambda$ and positive recurrence when $\gamma > \lambda$. This result is contained in a result from [Ly] giving recurrence and transience conditions for trees that are not necessarily Galton-Watson but that have this same transition vector that is a deterministic function of the number of children of the vertex.

Example: Suppose the tree is binary and has vertices of types A and B with probabilities p and $1-p$ respectively. Type A vertices have $\vec{p} = (1/3, 1/3, 1/3)$ and type B vertices have $\vec{p} = (3/5, 1/5, 1/5)$ so that type B vertices favor transitions to their parents (see figure 2). Then

$$\begin{aligned} m(\ln(r)) &= \inf_t \mathbf{E}(\phi/r)^t \\ &= \inf_t p(1/r)^t + (1-p)(1/3r)^t \end{aligned}$$

$$\text{which has a minimum at } t = \frac{\ln((1-p)\ln(3r)/p\ln(1/r))}{\ln(3)}.$$

Then

$$\begin{aligned} \lambda r m(\ln(r)) &= 2p(1/r)^{\{\ln[(1-p)\ln(3r)/p\ln(1/r)]/\ln(3)\}-1} \\ &+ \frac{2}{3}(1-p)(1/3r)^{\{\ln[(1-p)\ln(3r)/p\ln(1/r)]/\ln(3)\}-1}. \end{aligned} \tag{12}$$

For $p = 1/2$ the value $r = 1$ makes the first term at least 1 so the walk is transient according to (7) or (9). So the critical value for p is less than $1/2$. This can be seen directly by noting that $p = 1/2$ is the value at which vertices of type A percolate and for any $p > 1/2$ there is a subtree of type A vertices with branching number greater than 1. The walk restricted to this subtree is a simple random walk and clearly transient. Further discussion of the case $r = 1$ precedes the proof of lemma 4.10. Of course the critical value of p may be recovered precisely from (12) with the aid of a calculator.

4.4 Reduction to RWRE

In this section we study edge-reinforced random walk in order to prove the equivalence in lemma 4.7 below.

Fix a single node, v. It has parent v_0, and children v_i for some possibly empty set of i. Edges e_i connect v to v_i. When the edge-reinforced random walk first reaches v,

FIGURE 2

the edge weights must be $1+\Delta$ for e_0 and 1 for each other edge. If the walk returns later to v it must do so along the same edge by which it left. So the weight of one edge will increase by 2Δ while the others remain fixed. As long as the walk keeps returning to v, the weights of e_i increment in this fashion. It is easy to see that the sequence of edges by which the walk leaves v is an exchangeable sequence stopped at a random time. More precisely, let the (possibly finite)n sequence of times that the walk is at v be $t_1, t_2, \ldots$ and let X_i be the position of the walk at time $1+t_i$. Then for any finite sequence of vertices $W_1, \ldots, W_n$ and any permutation $\pi \in \mathcal{S}_n$,

$$\begin{aligned} & prob(X_1 = W_1, \ldots, X_n = W_n | \text{ at leat } n \text{ visits to } v) \\ = \; & prob(X_1 = W_{\pi(1)}, \ldots, X_n = W_{\pi(n)} | \text{ at leat } n \text{ visits to } v). \end{aligned}$$

We will see that this is in fact it is equivalent to Pólya's Urn with appropriate initial conditions.

Pólya's Urn contains balls of different colors. At each turn a ball is drawn and replaced along with D extra balls of the same color. Of course the probability of choosing a color is just the fraction of balls in the urn of that color. The mathematics still makes sense (although the mechanism does not) if D is allowed to be non-integral. Letting $D = 2\Delta$ and the initial numbers of each color be $1+\Delta$ for color 0 and 1 for each other color, gives the sequence of edges chosen at each visit to v. Combining theorem 2.2 with the remarks following theorem 2.1, we have

Lemma 4.6 (Multicolor Pólya's Urn) *Let the urn begin with w_i balls of color, $1 \le i \le k$. Then the sequence of draws is distributed as a mixture of sequences of i.i.d. draws with the common probability of choosing color i being a random variable, p_i. The vector $\vec{p}$ ranges over the unit simplex and has the Dirichlet distribution with parameters $w_1/D, \ldots, w_k/D$. In particular if $W = w_1 + \ldots + w_k$ then the density of p_i on $(0,1)$ is given by*

$$[\Gamma(W)/\Gamma(w_i/D)\Gamma(W - w_i/D)]\, x^{(w_i/D-1)}(1-x)^{(W-w_i/D-1)} \tag{13}$$

$\square$

A consequence of the acyclicity of the graph is that the sequence of edges chosen from v is (except for a random stopping time) independent of what happens on edges not incident to v. To clarify, take for example a vertex v with children $v_1, \ldots, v_k$. Let the sequence of times the walk is at v be $t_1, t_2, \ldots$ and let X_i be the position of the walk at time $1 + t_i$. Then the probability of a visit to v_j at time $1 + t_n$ given $t_n < \infty$, the shape of the tree and the σ-field of events upt to time t_n is equal to the probability of a visit to v_j given only that $t_n < \infty$ and the positions of the walk at times $1 + t_i$ for $i < n$, and both probabilities are given by

$$\frac{1 + 2\Delta(\text{number of } X_i \text{ equal to } v_j \text{ for } i < n)}{1 + k + (2n - 1)\Delta}.$$

But this is the same probability of picking a ball of color j from an urn that began

with $1 + \Delta$ balls of color 0 and one ball each of colors $1, \ldots, n$ and to which 2Δ balls have been added of each color $X_1, \ldots, X_n$. So we can model the edge-reinforced random walk by independent Pólya's Urns at each node, making the decisions about where to go from that node. The urns can be replaced in turn, according to the lemma, by random values $p_0(v)$, $p_1(v), \ldots, p_{M(v)}(v)$ chosen independently from the Dirichlet distribution with parameters $(1 + \Delta)/2\Delta, 1/2\Delta, \ldots, 1/2\Delta$. Conditional upon these choices, the walk is a Markov chain with transition probabilities *prob* $(v \to \text{par}(v)) = p_0(v)$ and *prob* $(v \to ci(v)) = p_i(v)$ for $1 \leq i \leq M(v)$. Another way of saying this is that the walk is governed by a transition matrix whose rows are independent Dirichlet. The following equivalence should now be clear.

Lemma 4.7 *Let $v_1, v_2, \ldots$ be the sequence of nodes visited by an RWRE on a Galton-Watson random tree whose transition matrix, conditioned on the shape of the tree, has rows whose nonzero elements are independent Dirichlet with parameters* $(1 + \Delta)/2\Delta$,

$1/2\Delta, \ldots, 1/2\Delta$. *The distribution of the random sequence $v_1, v_2, \ldots$ is the same as the distribution of sequences of nodes visited by an edge-reinforced random walk.* $\square$

For a binary tree, we give a formal construction of this which will be useful for the numerical calculation in equation (45) below. The formalisms for the general tree are equally routine. The beta density with parameters a and b is defined as

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}.$$

To construct the Dirichlet distribution out of betas, use the following.

Fact: If X is has beta distribution with parameters a and $b+c$ and Y is independent of X and has beta distribution with parameters b and c, then the vector $(X, (1-X)Y, (1-X)(1-Y))$ is Dirichlet with parameters a, b and c.

Let $\{A(v), B(v) : v \in \mathbf{T}\}$ be independent random variables with $A(\rho) = 0$, density for A for $v \neq \rho$ given by

$$[\Gamma((3+\Delta)/2\Delta)/\Gamma((1+\Delta)/2\Delta)\Gamma(1/\Delta)]\, x^{(1-\Delta)/2\Delta}(1-x)^{1/\Delta-1} \tag{14}$$

and density of B given by

$$[\Gamma(1/\Delta)/\Gamma(1/2\Delta)\Gamma(1/2\Delta)]\, x^{1/2\Delta-1}(1-x)^{1/2\Delta-1}. \tag{15}$$

So the variables $A(v)$ are i.i.d. betas with parameters $(1+\Delta)/2\Delta$ and $1/\Delta$ and the variables $B(v)$ are i.i.d. betas with parameters $1/2\Delta$ and $1/2\Delta$. Then the vector $(A(v)\,,\ (1-A(v))B(v)\,,\ (1-A(v))(1-B(v)))$ has the Dirichlet distribution with parameters $(1+\Delta)/2\Delta$, $1/2\Delta$ and $1/2\Delta$.

To construct the walk itself, let Z_i be i.i.d. uniform on (0,1) for $i = 1, 2, \ldots$, and let the random variable

$$\prod_{v \in \mathbf{T}} A(v) \times \prod_{v \in \mathbf{T}} B(v) \times \prod_{i \in \mathbf{N}} Z_i$$

be defined on some space Ω. For each $\omega \in \Omega$, generate a sequence of nodes recursively by $v_1(\omega) = \rho$ and $v_{n+1}(\omega) =$

$$\begin{cases} \mathrm{par}(v_n(\omega)) & \text{if} \quad Z_n(\omega) < A(v_n(\omega)) \\ c_1(v_n(\omega)) & \text{if} \quad A(v_n(\omega)) \leq Z_n(\omega) < A(v_n(\omega)) + (1 - A(v_n(\omega)))B(v_n(\omega)) \\ c_2(v_n(\omega)) & \quad\;\; \text{otherwise} \end{cases} .$$

The process $\omega \to (v_1(\omega), v_2(\omega), \ldots)$ is our RWRE. $A(v, \omega)$ and $B(v, \omega)$ are the random environment; conditional upon their values for all $v \in \mathbf{T}$, the Z_i determine a Markov random walk. [Unfortunately the distribution function for A does not vary pointwise monotonically with respect to Δ so it is difficult to compare the RWRE's for different values of Δ.]

We can now prove that the mean recurrence time is infinite even for $\Delta < 1$. Let R be the mean recurrence time given the values of A and B at all nodes so that the mean recurrence time is $\mathbf{E}(R)$. By looking at the subtree below the first node visited we get the equation $\mathbf{E}(R) = 2 + \mathbf{E}((1 - A)/A)\ \mathbf{E}(R)$. But for $0 \leq \Delta < 1$, we get $1 \leq \mathbf{E}(((1 - A)/A) < \infty$ by a calculation similar to (6.2) below, so $\mathbf{E}(R)$ must be infinite.

Proposition 4.8 *Let an RWRE on a Galton-Watson tree be i.i.d. in the sense of definition 4.4. Then the probability that the environment is recurrent conditioned on the tree being infinite is either 0 or 1.*

Proof: First consider the simpler case where the tree is binary. The environment is given by a set of independent random variables $\{A(v), B(v) : v \in \mathbf{T}\}$. Altering the values of A and B for finitely many v will not affect whether an environment is recurrent (since $A(v)$ and $B(v)$ are in the open interval $(0, 1)$ but may not equal 0 or 1). So recurrence is a tail property of the i.i.d. pairs $(A(v), B(v))$ given the number of children of each node. It follows from the 0-1 law for tails that the environment must be a.s. transient or a.s. recurrent. In other words, the mixing measure for the RWRE does not mix recurrence and transience, so the claim at the beginning of section 3 is established.

For general trees one must first condition on the trees being infinite. We use an argument due to H. Kesten. The process is recurrent iff the process restricted to each subtree with root v for $v \in \mathbf{T}_1$ is recurrent. So the recurrence probability is a fixed point of the offspring generating function. If it is less than 1 it is bounded by the extinction probability and hence equal to the extinction probability. So conditional upon non-extinction, the recurrence probability is either 1 or 0.

4.5 Transience of RWRE

In this section and the next, let $\mathbf{T}$ be a Galton-Watson tree and let the transition vectors $\vec{p}(v) = (p_1(v), \ldots, p_{M(v)}(v))$ be an i.i.d. RWRE as in definition 4.4. Let

$$A(v) = p_0(v) = prob(\text{transition from } v \text{ to } \text{par}(v)),$$

let

$$C(v) = p_i(\text{par}(v)) = prob(\text{transition from par}(v) \text{ to } v,$$

and let

$$\phi(v) = C(v)/A(\text{par}(v)). \tag{16}$$

In particular if we let

$$C(c_1(v)) = (1 - A(v))B(v) \text{ and} \tag{17}$$
$$C(c_2(v)) = (1 - A(v))(1 - B(v)) \tag{18}$$

then this agrees with the definition of ϕ in (5). The main results of this section are the transience criteria for RWRE, (7) and (9). We restate them here:

Theorem 4.9 *Let m be defined by equation (6). Suppose that for some $r \in (0,1]$*

$$\lambda r \, m(\ln(r)) > 1 \,. \tag{19}$$

Then the RWRE is transient.

The proof breaks into three pieces: lemma 4.10, which is a transience criterion for a single environment; Chernoff's identification of the rate function for large deviations; and a lemma on branching processes, providing the hypotheses for lemma 4.10.

Lemma 4.10 *Let $k \in \mathbf{N}, M \in \mathbf{R}^+, r \in (0,1]$ and $\delta > 0$ be fixed constants. For a set of nodes S let S_i denote $S \bigcap \mathbf{T}_{ik}$, the nodes of S at distance ik from the root. Suppose a nonempty set of nodes $S \in \mathbf{T}$ can be found such that for all v_0:*

$$[v \in S \text{ and } v_0 < v] \Rightarrow v \in S \tag{20}$$

$$v_0 \in S_1 \Rightarrow card\{v \in S_{i+1} : v_0 < v\} \geq r^k \tag{21}$$

For each branch segment $v_0 < v_1 < \ldots < v_k$ with $v_0 \in S_1$ and $v_k \in S_{i+1}$,

$$\sum_{1 \leq i \leq k} \ln(\phi(v_i)) \geq k \ln(r) + \delta \tag{22}$$

$$\phi(v)^{-1} \leq M \text{ for all } v \in S \,. \tag{23}$$

Then the environment is transient.

To see the intuition behind this lemma, suppose $r = 1$. Then (20) and (21) say that S contains at least one infinite branch. By (22) and (23), the $\liminf$ average of $\ln(\phi)$ along initial segments of any branch in S is at least δ/k. For any such branch, the Markov chain gotten by considering only moves along that branch will be transient;

this is because the sequence $\epsilon_i = \prod_{v<v_i} \phi(v)^{-1}$ is summable which is the standard test for transience in the one-dimensional case. But any environment containing a transient subtree is transient either because it wanders to infinity on the subtree or because it fails to return to the subtree infinitely often.

Proof of lemma 4.10: We find the appropriate martingale to generalize the standard test to the case $r < 1$. Define a function $s : \mathbf{T} \setminus \{\rho\} \to [0,1]$ by $s(v) = 0$ for $v \notin S$ and by

$$s(v) = \text{card}\{v' \in S_{i+1} : v \leq v'\} \;/\; \text{card}\{v' \in S_{i+1} : \text{par}(v) \leq v'\}$$

for $v \in \mathbf{T}_j$ with $ik < j \leq (i+j)k$. Clearly, the sum over i of $s(ci(v))$ is 1 for any $v \in S$. See figure 3 for an example of the function s. Now define $t : \mathbf{T} \to \mathbf{R}^+$ by $t(\rho) = 1$ and for $v \neq \rho$ by $t(v) = s(v)\phi(v)^{-1}t(\text{par }(v))$. Define $u : \mathbf{T} \to \mathbf{R}^+$ by $u(v) = \sum_{v' \leq v} t(v')$.

I claim that $u(v_i)$ is a bounded martingale for any $v_1 \neq \rho$, where $v_1, v_2, \ldots$ is a random walk on the given environment, stopped if it reaches ρ. To see if it is a martingale, just calculate

$$\begin{aligned}
\mathbf{E}(u(v_{i+1})|v_i) &= A(v_i)u(\text{par}(v_i)) + \sum_j C(cj(v_i))u(cj(v_i)) \\
&= u(v_i) + A(v_i)\left[-t(v_i) + \sum_j \phi(cj(v_i))t(cj(v_i))\right] \\
&= u(v_i) + A(v_i/t(v_i)\left[-1 + \sum_j s(cj(v_i))\right] = u(v_i)
\end{aligned} \tag{24}$$

for $v_i \in S$. For $v_i \notin S$ the result is true because $t(v_i) = 0$.

For boundedness, first consider the case $v \in S_i$. Find $v_0 \in S_{i-1}$ with $v_0 < v$. Then $\prod_{v_0 < v' \leq v} s(v')$ is a telescoping product and is at most r^k by (21). Also $\prod_{v_0 \leq v' \leq v} \phi(v)^{-1} \leq e^{-\delta}r^{-k}$ by (22). Then $t(v) \leq t(v_0)e^{-\delta}$ and by induction $t(v) \leq e^{-i\delta}$. Now for any $v \in S \bigcap \mathbf{T}_n$, apply (23) to see that $t(v)$ decreases at least geometrically in n. Therefore u is bounded on S. But for $v \notin S, u(v) = u(\text{par}(v))$ so u bounded on all of $\mathbf{T}$.

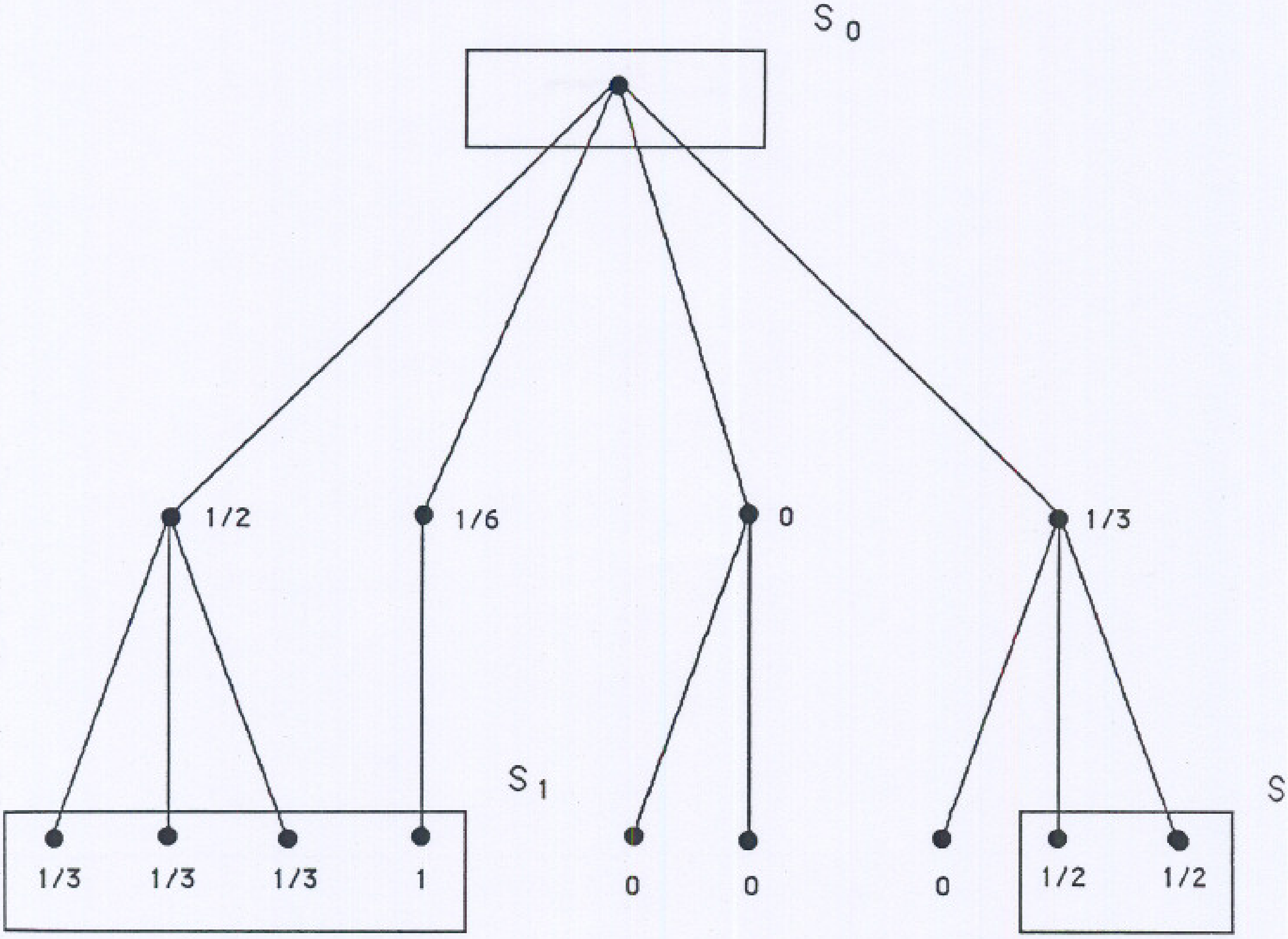


FIGURE 3

We conclude by the bounded martingale theorem that $u(v_i)$ converges a.s. to a limit u with $\mathbf{E}(u) = \mathbf{E}(u(v_1)) > 1$. Then prob$(u = 1) < 1$ so the walk stays away from the root with nonzero probability. □

The case (7) is easily disposed of using lemma 4.10 and the strong law of large numbers, so we assume for the remainder of this section that $\mathbf{E}(\ln(\phi)) < 0$. At this point we require Chernoff's estimates for the probabilities of large deviations.

Theorem 4.11 ([Ch] theorem 1 and lemma 6) *Let $S_n = X_1 + \ldots + X_n$ where the X_i are i.i.d. with common distribution function F. Define*

$$m(r, F) = \inf\{exp(-rt)\mathbf{E}(exp(tX_1)) : t \in \mathbf{R}\}. \tag{25}$$

Assume $r > \mathbf{E}(X_1) \geq -\infty$. Then

$$prob(S_n \geq nr) \leq m(r, F)^n \quad and \tag{26}$$

$$\lim_{n\to\infty} m_1^{-n} prob(S_n \geq nr) = \infty \text{ for any } m_1 < m(r, F). \tag{27}$$

Furthermore, $m(r, F)$ is continuous in r and strictly decreasing between $C = \mathbf{E}(X_1)$ and $D =$ essential $\sup(X_1)$ with $m(C, F) = 1$ and $m(D, F) = prob(X_1) = D$. □

We will apply this with $F = \Phi$, the distribution function for $\ln(\phi)$. With $m(r)$ denoting $m(r, \Phi)$, the notations in (25) and (6) agree. Roughly speaking, (27) tells us that under the condition (19), there are enough branches on which $\ln(\phi)$ averages more than $\ln(r)$ to made (20) - (23) possible. To make this into a proof we need some facts about branching processes.

The branching processes we will consider begin with a single ancestor. Each individual bears a random number of children which is i.i.d. and equals i with probability p_i. Let $f(x)$ be the generating function for the p_i's, with $f'(1) = M < \infty$, and assume $M > 1$ so the process is finite with some probability $b < 1$.

Lemma 4.12 *Pick any $K > 0, M_1 < M$. Then*

$$\lim_{n\to\infty} \text{prob}(\textit{size of the } n^{th} \textit{ generation } < KM_1^n) = b\,.$$

Proof: See [Ha] pps. 13-14 theorem 8.1 and remark 1. □

Say a branching process is d-infinite for $d \in \mathbf{N}$ if there is some nonempty subset of individuals such that each individual in the subset has at least d children in the subset. Say that a given individual has a d, n-subtree if $n = 0$ or the individual has at least d children each of whom has a $d, n-1$-subtree. Suppose that B is a branching process with generating function f and C is the process with generating function $f(r+(1-r)x)$, being identical to B except that births are aborted with probability r.

Lemma 4.13 *Suppose that for the process C, the probability of an individual having at least d children is at least $1-r$. Then the process B is d-infinite with probability at least $1-r$.*

Proof: We show by induction that the probability of any individual having a d, n-subtree is at least $1-r$. The case $n = 0$ is trivial. Now assume it is true for some arbitrary n. Then the probability of an individual having a $d, n+1$-subtree is just the probability of having at least d children, provided that the ones who will not have a d, n-subtree are aborted. By the induction hypothesis, children will be aborted with probability at most r, so by the hypothesis of the lemma, the probability of having a $d, n+1$-subtree is at least $1-r$.

Thus the probability of the initial ancestor having a d, n-subtree for all n is at least $1-r$. Since each node has only finitely many children, the process will be d-infinite in these cases.

□

For any branching process, B, let $B^{(k)}$ denote the process whose n^{th} generation is the nk^{th} generation of B, with the relation of parenthood in $B^{(k)}$ corresponding to ancestry in B.

Lemma 4.14 *For a branching process B, let f, b, M and M_1 be as in lemma 4.12 above. Then there is some $k \in \mathbf{N}$ such that*

$$\text{prob}(B^{(k)} \textit{ is } \lfloor M_1^k \rfloor\textit{-infinite}) \geq (1-b)/2. \tag{28}$$

Proof: By lemma 4.12 we can pick N large enough so that for all $i \geq N$

$$\textit{prob}(\text{size of } i^{th} \text{ generation of } B > 4M_1^i/(1-b)) > 3(1-b)/4. \tag{29}$$

By increasing N if necessary we can also assume that the following holds: given a population of size at least $4M_1^N/(1-b)$, each member of which is killed independently with probability $(1+b)/2$,

$$\textit{prob}(\text{at least } M_1^N \text{ of them survive}) > 3(1-b)/4. \tag{30}$$

Now let $k = N$ and apply lemma 4.13 to $B^{(k)}$ with the probability of abortion $= (1+b)/2$. Then

$$\begin{aligned}
&\textit{prob}(\text{having at least } M_1^N \text{ children in } B^{(N)} \text{ with abortion}) \\
\geq\ & 1 - \textit{prob}(\text{fewer than } 4M_1^N/(1-b) \text{ children in } B^{(N)} \text{ without abortion}) \\
&- \textit{prob}(\text{from at least } 4M_1^N/(1-b) \text{ conceptions fewer than } M_1^N \text{ are born}) \\
\\
\geq\ & 1 - (1-b)/4 - (1-b)/4 = (1-b)/2
\end{aligned}$$

So (28) follows from lemma 4.13. □

Proof of (9): Fix r with $\lambda r\ m(\ln(r)) > 1$. Fit in a few more constants: $m(\ln(r)) = (1+\delta_1)/\lambda r > (1+\delta_2)/\lambda r > (1+\delta_3)/\lambda r > (1+\delta_4)/\lambda r > 1/\lambda r$. Apply (27) of Chernoff's theorem with $m_1 = (1+\delta_2)/\lambda r < m(ln(r))$. For a vertex v picked uniformly from the n^{th} generation, (27) implies

$$\lim_{n\to\infty} \left(\frac{\lambda r}{(1+\delta_2)}\right)^n prob(\sum_{v'<v} \ln(\phi(v')) \geq n\ln(r)) = \infty$$

and in particular is eventually greater than λ. Now the expected number of nodes in the n^{th} generation is λ^n and if this number were independent of the values of ϕ it would immediately imply

$$\mathbf{E}(\mathrm{card}\{v \in \mathbf{T}_N : \sum_{v'<v} \ln(\phi(v')) > N\ln(r) + \delta_0\}) > \lambda((1+\delta_2)/r)^N \tag{31}$$

for sufficiently large N and small δ_0. In fact, the size of the n^{th} generation is independent enough of the values of ϕ to imply (31) if any smaller δ_3 is substituted for δ_2. The reason for this is that in any supercritical branching process with mean λ,

$$prob(|\mathbf{T}_n| \leq (\lambda(1+\delta_3)/(1+\delta_2)^n|\ \text{non-extinction}) \to 0 \text{ exponentially fast,}$$

and it is easy to see that conditioning on $|\mathbf{T}_n| \geq ((\lambda(1+\delta_3)/(1+\delta_2))^n$ can be made to have an arbitrarily small effect on $prob(\sum_{v'<v}\ln(\phi(v')) > N\ln(r)+\delta_0$. We can now pick M sufficiently large to amend (31) to

$$\begin{aligned}&\mathbf{E}(\mathrm{card}\{v \in \mathbf{T}_N : \sum_{v'<v} \ln(\phi(v')) > N\ln(r) + \delta_0 \text{ and } \phi(v')^{-1} < M \\ &\text{for all } v \leq v\}) > \lambda((1+\delta_3)/r)^N.\end{aligned} \tag{32}$$

Now define a branching process B with ρ as its initial ancestor, whose individuals are elements of $\mathbf{T}_0, \mathbf{T}_N, \mathbf{T}_{2N}, \ldots$ such that $v_0 \in \mathbf{T}_{iN}$ has $v \in \mathbf{T}_{(1+1)N}$ as a child iff $v_0 < v$ and $\sum_{v_0 \leq v' < v} \ln(\phi(v;)) \geq N\ln(r) + \delta_0$ and $\phi(v')^{-1} < M$ for all $v_0 \leq v' < v$ and v is the first child of par(v) that qualifies under these conditions. By lemma 4.14 there is a j such that $B^{(j)}$ is $\lfloor((1+\delta_4/r)^{jN}\rfloor$-infinite with nonzero probability. In fact j can be chosen large enough so that the expression in greatest-integer brackets is at least $(1/r)^{jN}$. Now the criterion given by lemma 4.10 applies with $k = jN$ to show that the probability of transience is nonzero. By the reasoning in section 4, this means the probability of transience, given an infinite tree, is 1 . $\square$

4.6 Recurrence of RWRE

The main result in this section is a proof of (8) above:

$$\text{If} \quad \mathbf{E}(\ln(\phi)) < 0 \text{ and } \sup\{\lambda r\, m(\ln(r)) : 0 < r \leq 1\} < 1 \text{ then the walk is a.s. positive recurrent;} \tag{8}$$

To prove (8) we calculate a stationary distribution. Sufficient conditions for a measure, μ, to be stationary are that for every v, i,

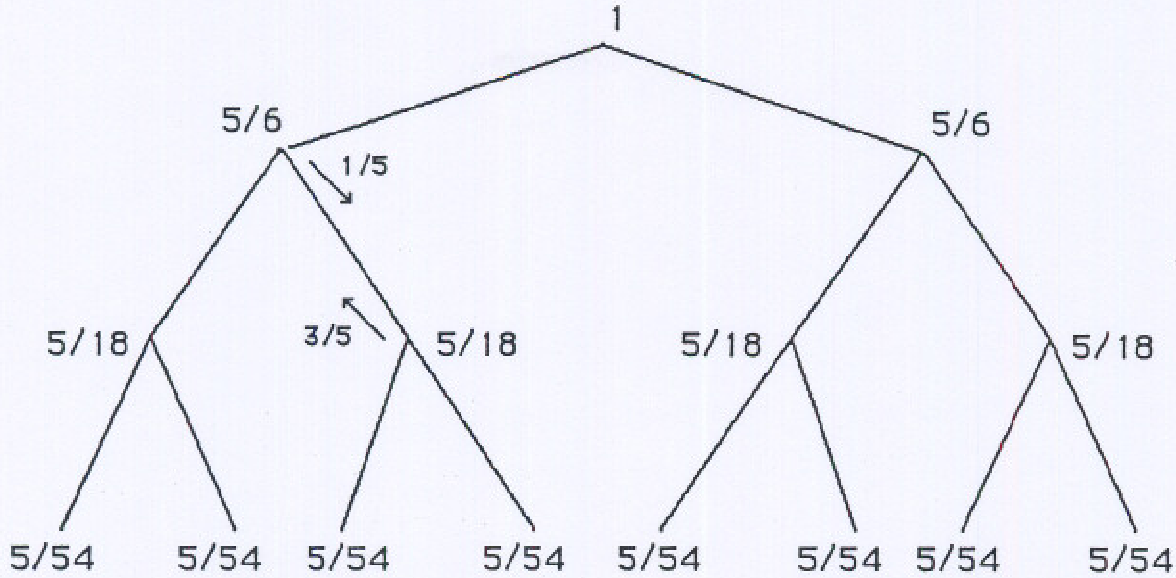
$$\mu(v)C(ci(v)) = \mu(ci(v))A(ci(v)). \tag{33}$$

If we let $\mu(\rho) = 1$ and for $v \neq \rho$ let

$$\mu(v) = A(v)^{-1} \prod_{\rho < v' \leq v} \phi(v') \tag{34}$$

then μ is stationary, satisfying (33). For example if a binary tree has $\vec{p}(v) = (3/5, 1/5, 1/5)$ for every v, then this measure gives $\mu(\rho) = 1$, $\mu(v) = 5/6$ for v in the first generation, $\mu(v) = 5/18$ for v in the second generation, $\mu(v) = 5/54$ for v in the third generation, and so forth (see figure 4). If $\mu(\mathbf{T}) < \infty$ then the walk is positive recurrent. The statement (10) follows immediately, since in this case $\mathbf{E}(\mu(\mathbf{T}))$ is finite so $\mu(\mathbf{T})$ is a.s. finite.

Roughly speaking, the reason $\mu(\mathbf{T})$ is finite under the hypotheses of (8) is that there are fewer than r^n nodes of measure r^n for each $r < 1$. This must be formulated



$$(1/5) \cdot (5/6) = (3/5) \cdot (5/18)$$

FIGURE 4

precisely and then integrated over $r \in [0,1]$. The methods are elementary, though in the case of lemma 4.16 a more elegant argument ought to be possible.

Let $f(v) = A(v)\mu(v)$. Note that $f(v)$ depends only on transition probabilities of nodes strictly above v.

Lemma 4.15 *Fix any $k \in (0,1]$. Assume $\mathbf{E}(\ln(\phi)) < \ln(r)$. Let $J_n = \{v \in \mathbf{T}_n : f(v) \geq r^n.\}$ Then prob(card(J_n) $\geq (\lambda km(\ln(r))^n)$ for infinitely many n) = 0.*

Proof: For each $v \in \mathbf{T}_n$, (26) gives $prob(f(v) \geq f^n) \leq m(\ln(r))^n$. So

$$\mathbf{E}(\mathrm{card}(J_n)) \leq (\lambda m(\ln(r)))^n \text{ and so} \tag{35}$$

$$\mathbf{E}(\sum \mathrm{card}(J_n)/(\lambda km(\ln(r)))^n) \text{ is finite.} \tag{36}$$

In the event that $\mathrm{card}(J_n) > (\lambda km(\ln(r)))^n$ infinitely often, the sum in (36) would be infinite; the event therefore has probability 0. □

Lemma 4.16 *Lemma 4.15 holds with μ in place of f.*

Proof: Let $G_n = \{v \in \mathbf{T}_n : \mu(v) \geq r^n.\}$ Suppose to the contrary that for some $a > 0$,

$$prob(\mathrm{card}(G_n) \geq (\lambda km(\ln(r)))^n \text{ infinitely often}) = a. \tag{37}$$

By continuity of m we can choose r_1 and k_1 so that $r > r_1 > 0$, $k > k_1 > 1$, $m(\ln(r_1)) < 1$ and $k_1 m(\ln(r_1)) \geq km(\ln(r))$. Then (37) holds with r_1 and k_1 in place of r and k. Pick k_2 so that $k_1 > k_2 > 1$ and $k_2 m(\ln(r_1)) = b < 1$ for some b. By lemma 4.15 with r_1 and k_2 in place of r and k, we can pick N_0 large enough so that

$$prob(\mathrm{card}\{v \in \mathbf{T}_n : f(v) \geq r_1^n\} \geq (\lambda k_2 m(\ln(r_1)))^n \text{ for some } n \geq N_0) < a/2. \tag{38}$$

By picking a larger N we can assume that $prob(C(v) \leq (r_1/r)^n) < \delta/L$ for any fixed δ, where L is a bound for $M(v)$. [This is the only place that the boundedness of M is used. Any weaker condition still implying the truth of this lemma can be substituted in (8).] We fix δ small enough so that

$$\begin{array}{l} \text{for any } n \geq \mathbf{N}_0 \text{ and any collection of individuals killed} \\ \text{independently with probability } \delta, \text{ the probability of the} \\ \text{fraction of survivors being at least } b \text{ is greater than } 1 - a/2\,. \end{array} \tag{39}$$

Now if the event in (37) occurs, pick the first $n \geq \mathbf{N}$ for which card $(G_n) \geq (\lambda k_1 m(\ln(r_1)))^n$. For any $v_0 \in G_n$ and any child v of v_0, $f(v) > r^{n+1}$ unless $C(v) \leq r_1^{n+1}/r^n < (r_1/r)^n$. Thus the event in (38) will hold for $n+1$ unless

$$\begin{aligned} &(\lambda k_1 m(\ln(r_1)))^n \text{card}\{v : \text{par}(v) \in G_n \text{ and } C(v) \leq (r_1/r)\} \div \\ &\text{card}\{v : \text{par}(v) \in G_n\}) \qquad\qquad \leq (\lambda k_2 m(\ln(r_1)))^{n+1}\,. \end{aligned} \tag{40}$$

But $(\lambda k_2 m(\ln(r_1)))^{n+1} < \lambda b(\lambda k_1 m(\ln(r_1)))^n$, so if (40) holds then

$$\text{card}(\{v : \text{par}(v) \in G_n \text{ and } C(v) \geq (r_1/r)^n\})/\lambda\text{card}(G_n) \leq b\,. \tag{41}$$

While the event $\text{par}(v) \in G_n$ is not independent of $C(v)$, it is easy to see that $prob(C(v) < x)$ can only decrease when conditioned on $\text{par}(v) \in G_n$. [Given the values of $A(v')$ and $B(v')$ for $\rho \leq v' < \text{par}(v)$, the indicator function of the event $\text{par}(v) \in G_n$ is a decreasing function of $A(\text{par}(v))$. Since $A(\text{par}(v))$ is independent of this σ-algebra, it follows that $\text{prob}(A(\text{par}(v)) < x)$ increases for any x when conditioned on $\text{par}(v) \in G_n$. Then $\text{prob}(C(v) < x)$ must decrease since $C(v) = (1 - A(\text{par}(v)))Y$ where Y is independent from all the above variables.] If we think of a node in G_n as being killed if the value of C at any of its children is less than $(r_1/r)^n$, then we can apply (39) to show that the probability of (41) is less than $a/2$. Thus the event in (37), having probability at least a, entails the disjunction of events in (38) and (41), each having probability less than $a/2$, a contradiction. $\square$

To finish proving (8) we use a compactness argument to estimate $\mu(\mathbf{T}_n)$. Let $\sup\{\lambda r\, m(\ln(r)) : r < 1\} = 1 - \delta_1$ and pick δ_2 and δ_3 with $1 - \delta_1 < 1 - \delta_2 < 1 - \delta_3 < 1$. Let $\mathcal{I}$ be the collection of intervals $\{(g(x), x) : x \in (0,1)\} \bigcup \{(g(1), 1]\}$ where g is any function such that

$$g(x) < x \text{ and } \lambda x m(\ln(g(x))) < 1 - \delta_3 \text{ for } x \in (0,1]. \tag{42}$$

Pick any b_0 with $0 < b_0 < 1/\lambda$. The elements of $\mathcal{I}$ cover $[b_0, 1]$ so, by compactness, pick a finite subcover J of $\mathcal{I}$. Let $(a_1, b_1), \ldots (a_k, b_k)$, $(a_{k+1}, 1]$ be the elements of J written in ascending order of a_i. Apply lemma (4.16) to each $(a, b) \in J$ with $r = a$ and $k = (1 - \delta_3)/(1 - \delta_2)$. Then there is almost surely some N such that for all $(a, b) \in J$ (including $a = 1$),

$$\operatorname{card}(\{v \in \mathbf{T}_n : \mu(v) \geq a^n\}) < [\lambda((1 - \delta_3)/(1 - \delta_2)) m(\ln(a))]^n \text{ for } n \geq N. \tag{43}$$

Then letting $a_0 = 0$, $b_0 = a_1$, $a_{k+1} = 1$ and assuming $n \geq N$ we get that

$$\begin{aligned}
\mu(\mathbf{T}_n) &\leq \sum_{0 \leq i \leq k+1} a_{i+1}^n \operatorname{card}\{v \in \mathbf{T}_n : \mu(v) \geq a_i\} \\
&\leq \lambda^n a_1^n + \sum_{1 \leq i \leq k} b_i^n [\lambda m(\ln(a_i))(1 - \delta_3)/(1 - \delta_2)]^n \text{ by (43)}, \\
&\leq (k + 10(1 - \delta_3))^n
\end{aligned}$$

by (42), so $\mu(\mathbf{T})$ is finite. □

4.7 Proof of theorem 4.3 and further questions

Applying (8) and (9) to equations (14) and (15), it remains only to calculate Δ_0. We first establish (11). Suppose $1 \leq \mathbf{E}(\ln(\phi)) \leq \infty$ and also that $m(0) < 1/\lambda$. Then for any $r \in (0, 1]$, (6) gives us

$$\lambda m(\ln(r)) = \inf\{\lambda r^{1-t} \mathbf{E}(\phi^t) : t \in \mathbf{R}\}. \tag{44}$$

By assumption this is less than 1 for $r = 1$. The infimum for $r = 1$ must occur at some $t \leq 1$ since $\mathbf{E}(\phi) \geq 1$ implies $\mathbf{E}(\phi^t) > (\mathbf{E}(\phi))^t > \mathbf{E}(\phi)$ for $t \geq 1$ by Jensen's inequality. But for positive r and $t \leq 1$, (44) is increasing in r, so the supremum of (44) over $r \in (0, 1]$ must be less than 1, establishing (11).

Now we calculate $m(0)$ as a function of Δ. Unravelling the definitions gives

$$\begin{aligned} m(0) &= \inf\{\mathbf{E}[B(1-A)/A]^t : t \in \mathbf{R}\} \\ &= \inf_t [\Gamma((3+\Delta)/2\Delta)\Gamma(1/2\Delta)/\Gamma((1+\Delta)/2\Delta)\Gamma(1/\Delta)\Gamma(1/2\Delta)\Gamma(1/2\Delta)] \\ &\quad \cdot \int\int [y(1-x)/x]^t x^{(1-\Delta)/2\Delta}(1-x)^{1/\Delta - 1} y^{1/2\Delta - 1}(1-y)^{1/2\Delta - 1} dx\, dy \\ &= \inf_t \Gamma((1+\Delta)/2\Delta - t)\Gamma(1/2\Delta + t)/\Gamma((1+\Delta)/2\Delta)\Gamma(1/2\Delta). \end{aligned} \tag{45}$$

Since $\log(\Gamma)$ is concave, the minimum is reached when $t = 1/4$, making the two factors in the numerator equal. For $\Delta \geq \Delta_0 \approx 4.29$, the expression (45) is less than 1/2. Using (44) and (45) with $t = 1$ gives $\mathbf{E}(\phi) = 1/(1+\Delta)$ for $\Delta > 1$ and $\mathbf{E}(\phi) = \infty$ for $\Delta \leq 1$. Also it is easy to see that $\mathbf{E}(\ln(\phi)) < 0$ for $\Delta > 1$ since for $\Delta = 1$ the distributions of A and $1 - A$ are identical. Then the conditions of (11) and (9) are satisfied for $\Delta > \Delta_0$, and (8) applies for $\Delta < \Delta_0$. □

Questions of edge-reinforced random walk on other graphs are still wide open. Diaconis originally asked me about the d-dimensional integer lattice $\mathbf{Z}^d$. I believe it is not even known whether there is a $\Delta > 0$ for which the edge-reinforced random walk on $\mathbf{Z}^2$ is recurrent!