Centers for random walks on trees

Andrew Beveridge
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

We consider two distinct centers which arise in measuring how quickly a random walk on a tree mixes. Lovász and Winkler [8] point out that stopping rules which "look where they are going" (rather than simply walking a fixed number of steps) can achieve a desired distribution exactly and efficiently. Considering an optimal stopping rule that reflects some aspect of mixing, we can use the expected length of this rule as a mixing measure. On trees, a number of these mixing measures identify particular nodes with central properties. In this context, we study a variety of natural notions of centrality. Each of these criteria identifies the barycenter of the tree as the "average" center and the newly defined focus as the "extremal" center.

Key words. Markov chain, random walk, stopping rule, tree, barycenter.

AMS Subject Classifications. 60510, 60G40, 05C05.

1 Introduction

What node is most central with respect to a random walk on a tree? Let the hitting time $H(i, j)$ be the expected time for a random walk starting at node i to get to node j. A natural definition for centrality is to require that the target node j minimize this hitting time for an appropriately chosen i. We consider two natural choices for this starting node. First, we identify the "average" center c by drawing i from the stationary distribution π:

$$\sum_i \pi_i H(i, c) = \min_j \sum_i \pi_i H(i, j). \tag{1}$$

Next we choose the worst possible starting node for each target j. Let j' be a j-*pessimal* node satisfying $H(j', j) = \max_i H(i, j)$. A target node a achieving

$$H(a', a) = \min_j H(j', j) = \min_j \max_i H(i, j) \tag{2}$$

is the "extremal" center of the tree.

These two notions of centrality are distinct. Let $B_{n,k}$ denote the broom graph consisting of a star of $n - k$ vertices centered at c and a path of length k also starting at c. If the nodes in the path are labeled $c = v_0, v_1, \ldots, v_k$, then simple calculations (see section 2 for methods) for $B_{10,7}$ show that node $c = v_0$ is its average center and node v_1 is its extremal center. In general, the barycenter satisfies equation (1), while the newly defined focus (or foci) of the tree satisfies equation (2). Moreover, we consider a variety of criteria for centrality with respect to random walks on trees, each of which characterizes an average center or an extremal center. The barycenter satisfies each average criterion and one of the foci of the tree (or both) satisfies by each extremal criterion.

For any undirected graph, Coppersmith, Tetali and Winkler [5] define the *central node* to be a node c for which $H(i, c) \leq H(c, i)$ for every node i. This definition of centrality falls into the average category since we are comparing c with all the other nodes in the graph. Indeed, we have

Proposition 1 *Let $G = (V, E)$ be a tree. The following statements for a node c are equivalent.*

    (a) The node c is a barycenter of the tree.

    (b) The node c satisfies $H(i, c) \leq H(c, i)$ for every node i.

    (c) $\sum_i \pi_i H(i, c) = \min_j \sum_i \pi_i H(i, j)$.

    (d) Every connected component $G' = (V', E')$ of $G \backslash \{c\}$ satisfies $\sum_{v \in V'} \pi(v) \leq 1/2$.

This equivalence has been known to the authors of [5].

Given an initial node i and a target distribution τ, we can follow an optimal stopping rule (see the next section for an exact definition) to halt a random walk starting at i so that the distribution of the final node is exactly τ. Denote the expected length of this optimal rule by $H(i, \tau)$. A number of parameterless mixing measures defined via stopping rules have been introduced and studied [1], [8], [9], [2], [3]. Among the most important measures are the mixing time $T_{\text{mix}} = \max_i H(i, \pi)$ and the reset time $T_{\text{reset}} = \sum_i \pi_i H(i, \pi)$. We interpret T_{mix} as the pessimal mixing time and T_{reset} as the average mixing time.

Since the barycenter is so closely related to average mixing, a natural question is how $H(c, \pi)$ compares with T_{reset}.

Proposition 2 $H(c, \pi) \leq 2T_{\text{reset}}$ *where c is a barycenter of the tree.*

However, $H(c, \pi)$ may be considerably smaller than T_{reset}. A rooted m-ary tree of depth r has its root c as its unique barycenter and unique focus. Fixing m and letting $r \to \infty$, the forget time (and hence the reset time) for this tree is $\Theta(m^{r-1})$ while the mixing time from the root c is $\Theta(r)$.

Theorem 3 *The mixing time from the root c of the m-ary tree of depth r is*

$$H(c, \pi) = \frac{(m+1)(m^r+1)}{(m-1)(m^r-1)}r - \frac{m^2 + 6m + 1}{2(m-1)^2}.$$

This exact mixing result, complementary to a result of Diaconis and Fill [6] is of independent interest.

In order to define the foci of the tree, we consider stopping rules for random walks. To each target distribution τ, we associate one node or two adjacent nodes, called the foci of τ.

Definition. A *focus u* of a distribution τ on the tree G is a node for which the rule "take one step from u and then follow an optimal rule from this random neighbor of u to τ" is *not* optimal. The *focus (foci) of the tree* is the focus (foci) for π.

For example, consider the path on 3 nodes and take our target to be $\pi = (1/4, 1/2, 1/4)$. Starting from the center, the rule "with probability $1/2$ take one step, otherwise stay put" is a stopping rule which achieves π. Results in the next section show that this rule is optimal, with expected length $1/2$. However, the rule "take one step and then follow an optimal rule to π" is clearly not optimal since this rule has expected length greater than 1. On the other hand, starting from an endpoint, the rule "take one step and then follow an optimal rule from the center to π" turns out to be an optimal rule, so the center is the unique focus for π in the 3-path. An analogous argument shows that the two internal nodes of the 4-path are the foci for π. A similar phenomenon holds in general.

Theorem 4 *Every distribution τ on a tree has either one focus or two adjacent foci.*

A key observation is that for any node i, the rule "walk from i to the nearest focus and then follow an optimal rule from that focus to τ" is an optimal rule from i to τ. In other words, the foci of τ are central with respect to all walks from nodes to τ. Naturally, we define the foci of the tree G to be the foci of the stationary distribution π. We call the tree *focal* if there it has a single focus, and *bifocal* if there are two foci. These foci are the extremal center of G: one of them always achieves the minimum of (2).

Proposition 5 *If the node a minimizes the pessimal hitting time $H(a', a) = \min_i H(i', i)$ then a is a focus of G.*

A focus which minimizes the pessimal hitting time is called a *primary focus* and we denote it by a. Otherwise, the node is a *secondary focus*, denoted b.

Another important parameterless mixing measure is the forget time $T_{\text{forget}} = \min_\tau \max_i H(i, \tau)$ where the minimum is taken over all target distributions. We interpret this quantity as the minimum expected time to "forget" the node we started from by following an optimal rule to some distribution. In spite of its rather unorthodox definition, the forget time is intimately connected to the mixing time and the reset time. For an undirected graph, $T_{\text{reset}} = T_{\text{forget}}$ (see [9]) and they are within a factor of 4 of T_{mix} (see [2]).

For any graph, Lovász and Winkler [9] show that there is a unique distribution μ achieving the forget time. This distribution μ is central in an extremal sense: μ minimizes the expected length of a rule starting from the worst possible node. For a tree, μ is concentrated on the foci of G:

Proposition 6 *If the node a is the unique focus of $G = (V, E)$ then μ is the singleton distribution on the focus a. If the adjacent nodes a and b are the foci of G then*

$$\mu_i = \begin{cases} (H(b', b) - H(a', b))/2|E|, & i = a \\ (H(a', a) - H(b', a))/2|E|, & i = b \\ 0, & otherwise. \end{cases}$$

where $H(i', i) = \max_j H(j, i)$.

Another mixing measure with central properties is $T_{\text{bestmix}} = \min_i H(i, \pi)$. The node achieving T_{bestmix} is the best possible starting node for achieving the stationary distribution. This formulation is dual in some sense to that of equation (1). As expected, the foci of the tree are central for this extremal problem.

Theorem 7 *The quantity $T_{\text{bestmix}} = \min_i H(i, \pi)$ is achieved by a focus of the tree. Specifically, if $H(a', b) < H(b', a)$ then a uniquely achieves T_{bestmix}, if $H(a', b) > H(b', a)$ then b uniquely achieves T_{bestmix} and if $H(a', b) = H(b', a)$ then T_{bestmix} is achieved by both a and b.*

Consider another mixing measure similar to the forget time. The start-independent time of a distribution σ is $T_{\text{si}}(\sigma) = \min_\tau \sum_i \sigma_i H(i, \tau)$ where the minimum is taken over all target distributions. For a walk started from σ, $T_{\text{si}}(\sigma)$ is the minimum expected time to obtain a

sample (from some distribution) that is independent of the initial node of the walk (which was drawn from σ). We may interpret $T_{\mathrm{si}}(\sigma)$ as the fastest way to "forget" that we started our walk from a node drawn from σ. A natural choice for our initial distribution is the stationary distribution. For a tree, the target distribution achieving $T_{\mathrm{si}}(\pi)$ is central in an average sense and indeed the optimal target is concentrated on the barycenter.

Proposition 8 *If c is a barycenter of the tree G then*

$$T_{\mathrm{si}}(\pi) = \min_\tau \sum_k \pi_k H(k,\tau) = \sum_k \pi_k H(k,c) = H(\pi,c).$$

By definition, $T_{\mathrm{si}}(\pi) \leq T_{\mathrm{reset}}$ and there are trees for which $T_{\mathrm{si}}(\pi)$ is arbitrarily small when compared with T_{reset}.

We define the start-independent time of a graph to be $T_{\mathrm{si}} = \max_\sigma T_{\mathrm{si}}(\sigma)$ where the maximum is taken over all initial distributions. For a tree, the target distribution achieving T_{si} is central in an extremal sense, and indeed T_{si} can be achieved by taking either focus as a target. Furthermore, $T_{\mathrm{si}} = T_{\mathrm{forget}}$.

Theorem 9 *For a tree G, we have $T_{\mathrm{si}} = T_{\mathrm{forget}}$. Moreover, if G has a unique focus a, then there exists a distribution ϕ such that $T_{\mathrm{si}} = H(\phi, a)$. If G has two foci a and b then there exists a distribution ϕ such that $T_{\mathrm{si}} = H(\phi, a) = H(\phi, b)$.*

It is an open question how $T_{\mathrm{si}}(\pi)$ and T_{si} compare to the other mixing measures for general graphs.

2   Preliminaries

Random walks. Given an undirected, connected graph $G = (V, E)$ a random walk on G is a sequence of nodes $(w_0, w_1, \ldots, w_t, \ldots)$ such that the node w_t at time t is chosen uniformly from the neighbors of w_{t-1}. For non-bipartite G, as t tends to infinity the distribution of the tth node tends to the so called stationary distribution π where $\pi_i = d(i)/2|E|$ and $d(i)$ is the degree of i. For bipartite G, we have convergence if we consider a "lazy walk" in which at each step we stay at the current node with probability $1/2$.

We define $M = \{p_{ij}\}$ to be the matrix of transition probabilities, so $p_{ij} = 1/d(i)$ if $ij \in E$ and $p_{ij} = 0$ otherwise. For two nodes i, j, the *hitting time* $H(i, j)$ is the expected length of a walk from i to j. The expected number of steps before a walk started at i returns to i is

$$\mathrm{Ret}(i) = \frac{1}{\pi_i}. \tag{3}$$

Hitting times for trees. We give a hitting time formula for trees that is equivalent to those found in [4] and [10]. Given neighboring nodes i and j, let V_i be the nodes in connected component of $G\backslash\{j\}$ containing i. Then

$$H(i,j) = \sum_{k \in V_i} d(k) = 2|E| \sum_{k \in V_i} \pi_k = 2|E|\,\pi(V_i) \tag{4}$$

where $\pi(S) = \sum_{k \in S} \pi(k)$ for $S \subset V$. Indeed, let G' be the induced subgraph on the nodes $V_i \cup \{j\}$. Let $d'(k)$ be the G'-degree of k and $H_{G'}(i,j)$ be the hitting time from i to j for this graph. By (3), $H(i,j) = H_{G'}(i,j) = \text{Ret}_{G'}(j) - 1 = \sum_{k \in G'} d'(k) - 1 = \sum_{k \in V_i} d(k)$.

If i and j are neighbors then equation (4) immediately gives

$$H(i,j) + H(j,i) = 2|E|. \tag{5}$$

Furthermore, we can determine a hitting time formula for the general case. Define $\ell(i,k;j) = \frac{1}{2}(d(i,j) + d(k,j) - d(i,k))$, the length of the intersection of the (i,j)-path and the (k,j)-path. This function is symmetric in i and k and is zero if and only if $i = j$, $k = j$, or the nodes i and k are in different connected components of $G \setminus \{j\}$. Assume $d(i,j) = r$ and the (i,j)-path is given by $(i = i_0, i_1, i_2, \ldots, i_r = j)$. Using (4) and $\ell(i,k;j)$ yields

$$H(i,j) = \sum_{t=0}^{r-1} H(i_t, i_{t+1}) = \sum_k \ell(i,k;j)d(k). \tag{6}$$

Stopping rules. We briefly summarize some results of Lovász and Winkler [8]. Let V^* be the space of finite walks on V, i.e. the set of finite strings $w = (w_0, w_1, w_2, \ldots, w_t)$, $w_i \in V$ and w_i adjacent to w_{i-1}. For a given initial distribution σ, the probability of w being the walk after t steps is

$$\Pr(w) = \sigma_{w_0} \prod_{i=0}^{t-1} p_{w_i, w_{i+1}}.$$

A *stopping rule* Γ is a map from V^* to $[0,1]$ such that $\Gamma(w)$ is the probability of continuing given that w is the walk so far observed. We assume that with probability 1 the rule stops the walk in a finite number of steps.

Given another distribution τ on V, the *access time* $H(\sigma, \tau)$ is the minimum expected length of a stopping rule Γ that produces τ when started at σ. We say Γ is *optimal* if it achieves this minimum. Optimal stopping rules exist for any pair σ, τ of distributions and the access time $H(\sigma, \tau)$ has many useful algebraic properties. When σ and τ are concentrated on nodes i and j respectively (we write $\sigma = i, \tau = j$), the access time $H(i,j)$ is the hitting time from i to j. Clearly, $H(\sigma, j) = \sum_i \sigma_i H(i,j)$ and $H(\sigma, \tau) \leq \sum_i \sigma_i H(i, \tau)$.

6

Given a stopping rule Γ from σ to τ, for each $i \in S$ the *exit frequency* $x_i(\Gamma)$ is the expected number of times the walk leaves node i before halting. Exit frequencies are fundamental to virtually all access time results. A key observation, due to Pitman [12], is the "conservation equation"

$$\sum_i p_{ij} x_i(\Gamma) - x_j(\Gamma) = \tau_j - \sigma_j. \tag{7}$$

It follows that the exit frequencies for two rules from σ to τ differ by $K\pi_i$ where K is the difference between the expected lengths of these rules. Hence the distributions σ and τ uniquely determine the exit frequencies for a mean optimal stopping rule between them and we denote these optimal exit frequencies by $x_i(\sigma, \tau)$. Moreover, a stopping rule Γ is mean-optimal if and only if there exists a *halting state* k such that $x_k(\Gamma) = 0$.

Any three distributions ρ, σ and τ satisfy the "triangle inequality"

$$H(\rho, \tau) \leq H(\rho, \sigma) + H(\sigma, \tau) \tag{8}$$

with equality holding if and only if there is a k that is a halting state from ρ to σ and also a halting state from σ to τ. In particular, $H(\sigma, j) \leq H(\sigma, \tau) + H(\tau, j)$ and equality holds if and only if j is a halting state for an optimal rule from σ to τ. Hence

$$H(\sigma, \tau) = \max_j (H(\sigma, j) - H(\tau, j)). \tag{9}$$

In the case $\sigma = i$ and $\tau = \pi$, the node j is halting for i if and only if $H(j, i) = \max_k H(k, i)$. Let i' be such an *i-pessimal* node. We have another formula for the access time from i to π (which follows from (9) and (13) below):

$$H(i, \pi) = H(i', i) - H(\pi, i). \tag{10}$$

Mixing measures. Stopping rules provide a number of parameterless mixing measures. We define the *mixing time* T_{mix} to be the expected length of an optimal mixing rule starting from the worst initial node: $T_{\text{mix}} = \max_i H(i, \pi)$. A node achieving this maximum is called *mixing pessimal*. The *forget time* T_{forget} is the smallest t such that there exists a distribution μ such that for every starting node, the expected time to attain μ via an optimal rule is at most t: $T_{\text{forget}} = \min_\tau \max_i H(i, \tau)$. The unique distribution μ achieving T_{forget} is called the forget distribution and is given by

$$\mu_i = \pi_i \left(1 + \sum_j p_{ij} H(j, \pi) - H(i, \pi) \right). \tag{11}$$

Furthermore, if a node is mixing pessimal then it is also pessimal for μ and every mixing pessimal node is a halting state for an optimal rule from μ to π.

The *reset time* $T_{\text{reset}} = \sum_i \pi_i H(i, \pi)$ can be viewed as an average mixing time. One of the main results of [9] is the nontrivial equality $T_{\text{forget}} = T_{\text{reset}}$ for a random walk on an undirected graph. Moreover, for an undirected graph we have $T_{\text{forget}} \leq T_{\text{mix}} \leq 4T_{\text{forget}}$ (see [2]).

Start-independence. The following independence condition arises naturally in applications of random walks. Let Γ be a stopping rule from σ to τ and let $w_0, w_1, \ldots, w_T$ be a walk halted by Γ at time T. The *support* of σ, denoted S_σ, is the the set of nodes i such that $\sigma_i > 0$. We associate a conditional distribution $\tau^{(i)}$ to each $i \in S_\sigma$ given by $\tau_k^{(i)} = \Pr\{w_T = k | w_0 = i\}$. In other words, $\tau_k^{(i)}$ is the probability that Γ stops the walk at k given that the walk started at i (which was drawn from σ). Clearly $\sum_{S_\sigma} \sigma_i \tau^{(i)} = \tau$ and we call the set $\{\tau^{(i)}\}_{S_\sigma}$ the Γ-*decomposition* of τ.

The rule Γ is *start-independent* if $\tau^{(i)} = \tau$ for all $i \in S_\sigma$. The node at which a start-independent rule halts is independent of the initial node. Start-independent rules always exist: the rule "draw w_0 from σ and walk optimally from w_0 to τ" is a start-independent rule of expected length $\sum_i \sigma_i H(i, \tau)$.

While start-independent rules are rarely optimal, they arise naturally in applications requiring multiple independent samples from the stationary distribution of some state space. We obtain these samples by following an optimal mixing rule, accepting the current state, and then starting a new optimal mixing walk from this state. In this setting, T_{reset} is the expected length of a minimal start-independent rule from π to π. (See [3] for an extremal result concerning start-independent rules whose initial and target distributions are identical.)

We define the start-independent time of a distribution σ to be the minimum expected length of a start-independent rule with initial distribution σ:

$$T_{\text{si}}(\sigma) = \min_\tau \sum_i \sigma_i H(i, \tau)$$

A quantity of natural interest is $T_{\text{si}}(\pi)$, the start-independent time for the stationary distribution. We would also like to determine the extremal behavior of $T_{\text{si}}(\sigma)$. The start-independent time of any singleton distribution is zero, so only the maximum case is nontrivial. We define the start-independent time of the graph to be

$$T_{\text{si}} = \max_\sigma T_{\text{si}}(\sigma) = \max_\sigma \min_\tau \sum_i \sigma_i H(i, \tau).$$

Centers. There are two classical centers for trees. The node achieving $\min_i \max_j d(i, j)$,

8

where $d(i,j)$ is the length of the unique path between i and j, is the *center* of the tree G (or bicenter if there are two adjacent nodes achieving this minimum). In other words, the distance to the furthest node from the center is minimal among all nodes of the tree G. This vertex does not appear to have any central properties with respect to random walks. The *barycenter* is the node (or the two adjacent nodes) which achieves $\min_i \sum_j d(i,j)$. A barycenter minimizes the total distance to all other nodes.

For any undirected graph G (not necessarily a tree), Coppersmith, Tetali and Winkler [5] define a *central node* c of G to be a node which satisfies $H(i,c) \leq H(c,i)$ for all i. The existence of such a node for an undirected graph follows from their cycle reversing identity

$$H(i,j) + H(j,k) + H(k,i) = H(i,k) + H(k,j) + H(j,i). \tag{12}$$

The equivalence of (a) and (b) in proposition 1 follows from the "random target identity" (see [7] for example) : the sum $\sum_j \pi_j H(i,j)$ is independent of the initial node i. Multiplying (12) by π_k, summing over all k and using this identity gives

$$H(\pi,i) + H(i,j) = H(\pi,j) + H(j,i) \tag{13}$$

for any nodes i,j. Hence $H(i,c) \leq H(c,i)$ for all i if and only if $H(\pi,c) \leq H(\pi,i)$ for all i.

3   The barycenter of the tree

For any $S \subset V$ and a distribution τ on V, let $\tau(S) = \sum_{k \in S} \tau_k$ and $\bar{S} = V \backslash S$.

Proof of proposition 1. By equation (13), (b) and (c) are equivalent.

We show that (c) and (d) are equivalent. Assume $\pi(V') \leq 1/2$ for each component $G' = (V', E')$ of $G \backslash \{c\}$. For $i \in V'$, $H(i,c) \leq d(i,c) \sum_{V'} d(k) \leq d(c,i) \sum_{\bar{V}'} d(k) \leq H(c,i)$ by (6), so c is the central node. Now assume that c is the central node and that $\pi(V') > 1/2$. If c_0 is the unique neighbor of c in V' then $H(c_0,c) = \sum_{V'} d(k) > \sum_{\bar{V}'} d(k) = H(c,c_0)$, a contradiction. If $\pi(V') = 1/2$ for some component G' of $G \backslash \{c\}$ then the neighbor of c in V' is also a central node. Finally we prove the equivalence of (a) and (d). Let j be a node such that some component G' of $G \backslash \{j\}$ has $\pi(V') > 1/2$. Let j_0 be the unique neighbor of j in V'. We have $\sum_i d(i,j_0) = \sum_i d(i,j) - |V'| + |\bar{V}'| < \sum_i d(i,j)$, so j is not the barycenter of the tree.   □

The barycenter is the average center for random walks on trees, so it is natural to compare $H(c,\pi)$ and $T_{\text{reset}} = T_{\text{forget}}$. Mixing from the barycenter never takes more than twice as long as the average mixing time.

Proof of proposition 2. Let u be the unique neighbor of c on the path from c to a c-pessimal node c', and let G' be the connected component of $G \setminus \{u\}$ containing c. By proposition 1, $\pi(G') \geq 1/2$. For any node $i \in V(G')$, we have $H(i, \pi) = H(i, c) + H(c, \pi)$ since c' will also be i-pessimal. Therefore

$$T_{\text{reset}} \geq \sum_{i \in G'} \pi_i H(i, \pi) = \sum_{i \in G'} \pi_i \left(H(i, c) + H(c, \pi) \right) \geq \pi(G') H(c, \pi) \geq \frac{1}{2} H(c, \pi).$$

$\square$

There are trees for which $H(c, \pi) > T_{\text{reset}}$. Consider a broom graph $B_{k,k}$ where k is divisible by 4, with path nodes $(c = v_0, v_1, \ldots v_k)$. Some simple calculations using (6) show that $v_{k/4}$ and $v_{k/4+1}$ are the foci of $B_{k,k}$ and therefore the forget time (and hence the reset time) is not greater than $(1 + 3k/4)^2$. Using (6) and (10), the expected time to mix from the barycenter is $H(c, \pi) = H(c, c') - H(\pi, c') = (4k^2 - 1)/6$ which is strictly greater than $(1 + 3k/4)^2$ for large k. Of course, $(4k^2 - 1)/6 \leq 2(3k/4)^2 \leq 2T_{\text{reset}}$ as stipulated by proposition 2.

On the other hand, $H(c, \pi)$ may be markedly smaller than the forget time (and hence the reset time) of the tree. Consider an m-ary tree of depth r with root c. Of course c is the center, the barycenter and the focus of this tree. We adopt the following notation: $S_k = \{i | d(i, c) = k\}$ is the set of all nodes at level k. Let $c = i_0, i_1, \ldots, i_r$ be a path from c to a leaf i_r. The expected behavior of the walk at a node only depends on the level of the node, so we may use i_k as a representative for all nodes in S_k. A node is halting for this mixing walk if and only if it lies in S_r. We explicitly calculate $H(c, \pi) = H(i_r, c) - H(\pi, c)$ as per equation (10). Counting the degrees level-wise, the total number of edges in an m-ary tree of depth r is

$$\frac{1}{2}(m + m(m+1) + m^2(m+1) + \cdots + m^{r-1}(m-1) + m^r) = \frac{m(m^r - 1)}{m - 1}.$$

Proof of theorem 3. We start by showing that

$$H(i_r, i_{r-s}) = s + \frac{2m}{(m-1)^2}(m^s - sm + s - 1). \tag{14}$$

We partition G into sets $T_k = \{j | \ell(i_r, j; i_{r-s}) = k\}$ so that

$$H(i_r, i_{r-s}) = \sum_{k=1}^{s} k \sum_{j \in T_k} d(j)$$

by (6). We have $T_s = \{i_r\}$ and for $2 \leq k \leq r - 1$, T_k consists of the node i_k connected to $m - 1$ copies of m-ary trees of depth $r - k - 1$. Hence,

$$\sum_{j \in T_k} d(j) = (m + 1) + (m - 1) \left(1 + \frac{2m(m^{s-k} - 1)}{m - 1} \right) = 2m^{s-k+1}$$

10

for $2 \le k \le r - 1$ so that

$$
\begin{aligned}
H(i_r, i_{r-s}) &= s + \sum_1^{s-1} k(2m^{s-k}) = s + 2\sum_1^{s-1}(s-j)m^j \\
&= s + 2s\sum_1^{s-1} m^j - 2\sum_1^{s-1} jm^j = s + 2m\frac{m^s - sm + s - 1}{(m-1)^2}.
\end{aligned}
$$

Now, we denote $\pi(S_k) = \sum_{S_k} \pi_j$. By equation (10),

$$
\begin{aligned}
H(c, \pi) &= H(i_r, c) - H(\pi, c) = H(i_r, c) - \sum_j \pi_j H(j, c) \\
&= H(i_r, c) - \sum_{k=0}^r \pi(S_k) H(i_k, c) = \sum_{k=0}^r \pi(S_k) H(i_r, i_k) \\
&= \frac{m-1}{2m(m^r - 1)}\left(mH(i_r, i_0) + \sum_{k=1}^{r-1} m^k(m+1)H(i_r, i_{r-(r-k)}) + m^r H(i_r, i_r)\right).
\end{aligned}
$$

Using (14) and simplifying (we omit the details) yields the theorem. $\qquad\square$

The forget time for the m-ary tree of depth r is $H(i_r, c) = H(i_r, i_0) = \Theta(m^{r-1})$. Theorem 3 shows that $H(c, \pi) = \Theta(r)$. We compare the asymptotic behavior of this exact mixing result with a similar approximate mixing result of Diaconis and Fill [6]. Let σ_t denote the distribution achieved by walking t steps from the root c. Diaconis and Fill (example 4.20 of [6]) show that for a fixed m, as $r \to \infty$, $||\sigma_t - \pi||$ becomes small after $r(m+1)/(m-1) + \alpha r^{1/2}$ steps with α large. Theorem 3 shows that these two two mixing measures are asymptotically equivalent.

4 The foci of a distribution

The node k is τ-halting for i if $x_k(i, \tau) = 0$. Two nodes i, j have a common halting state for τ when there exists a node k such that $x_k(i, \tau) = 0$ and $x_k(j, \tau) = 0$. Recall that a focus u of a distribution τ on the tree G to be a node for which the rule "take one step from u and then follow an optimal rule from this random neighbor of u to τ" is not optimal, i.e. $H(u, \tau) < 1 + \sum_i p_{ui} H(i, \tau)$. This is equivalent to saying that there is no node that is simultaneously τ-halting for u and all of its neighbors.

For example, the focus for the singleton distribution $\tau = u$ is the node u. Considering mixing walks, equation (10) shows that k is a π-halting state for i if and only if $H(k, i) = \max_j H(j, i) = H(i', i)$. Hence for a path of even length the unique center is the only π-focus, and for a path of odd length, the two central nodes are the π-foci. Also, the center of a star graph is the only π-focus.

Theorem 4 states that for any tree G and any distribution τ, every τ has one focus or two adjacent foci. Fixing τ, let i^* denote a halting state for an optimal stopping rule from i to τ.

Proof of theorem 4. The case when τ is a singleton is trivial, so assume τ is not a singleton.

We claim that if i^* is a τ-halting state for i and j and i^* are in different components of $G\backslash\{i\}$ then i^* is a τ-halting state for j. We are guaranteed that i is on the unique (j, i^*)-path, so the rule "walk from j to i and then follow and optimal rule from i to τ" has i^* as a halting state and hence this composite rule from j to τ is optimal.

Next, consider the path $(j_1, \ldots, i_1, i_2, \ldots, j_2)$ in the tree G. We claim that the nodes j_1, j_2 cannot both be τ-halting states for each of the nodes i_1, i_2. Otherwise, equation (9) yields $H(i_k, \tau) = H(i_k, j_1) - H(\tau, j_1) = H(i_k, j_2) - H(\tau, j_2)$ for $k = 1, 2$ and therefore $-H(i_2, i_1) = H(i_1, j_1) - H(i_2, j_1) = H(i_1, \tau) - H(i_2, \tau) = H(i_1, j_2) - H(i_2, j_2) = H(i_1, i_2)$, a contradiction. Now there are two cases to consider. For two adjacent nodes u, v, let $V_u = \{k \,|\, d(u, k) < d(v, k)\}$ and $V_v = \{k \,|\, d(v, k) < d(u, k)\} = \bar{V}_u$.

CASE 1: There exists an edge uv such that u and v do not share a halting state for τ. Note that $u^* \in V_v$ and $v^* \in V_u$. Consider a set of optimal rules from the singletons to τ. As shown above, every node in V_u has u^* as a halting state, and u^* is not a halting state for any node in V_v. Similarly v^* is halting for all of V_v and is not halting for all of V_u. Hence u and v are the only foci of τ.

CASE 2: Every neighboring pair of nodes share a τ-halting state. Since τ is not a singleton, there exists a path of the form $(u^*, \ldots, i, u, \ldots, i^*)$ where u^* is a halting state for u but not for i, and u separates i from all of its τ-halting states. If u is not a focus, then the neighbors of u have a common halting state j^*. Let $j \neq i$ be the neighbor of u on the (u, j^*)-path. The path $(u^*, \ldots, u, j, \ldots, j^*)$ is of the form forbidden above, a contradiction. So u must be a focus and u is unique. Indeed, u shares a τ-halting state with each of its neighbors, so u must have a halting state in at least two components of $G\backslash\{u\}$. If there were another focus v, this would again imply the existence of a path forbidden above. $\square$

When τ has two foci u and v, we will continue to use the notation $V_u = \{k \,|\, d(u, k) < d(v, k)\}$ and $V_v = \{k \,|\, d(v, k) < d(u, k)\} = \bar{V}_u$. The following corollary is immediate from the proof.

Corollary 10 *If τ has a unique focus u then $H(i, \tau) = H(i, u) + H(u, \tau)$ for all i. If τ has two foci u, v then for $i \in V_u$, $H(i, \tau) = H(i, u) + H(u, \tau)$ and for $i \in V_v$, $H(i, \tau) = H(i, v) + H(v, \tau)$.*

Proof of proposition 5. Recall that a halting state i' for an optimal rule from i to π must

be i-pessimal (that is $H(i', i) = \max_j H(j, i)$). The node a is not a focus for π if and only if a' is halting for every neighbor of a. Let i be the neighbor of a on the (a, a')-path. The node a' is i-pessimal and $H(i, a') < H(a, a')$, a contradiction. $\qquad\square$

We say that this node a is a *primary focus* of G. We call G *focal* if π has a unique focus. Otherwise we call G *bifocal* and we denote the other focus by b. A bifocal tree may have two primary foci (for example, a path of odd length), but typically $H(b', b) > H(a', a)$ and we refer to b as the *secondary focus* of G. This new notion of center does not coincide with either the center or the barycenter of the tree: a few simple calculations for the broom graph $B_{10,7}$ shows that none of these nodes coincide.

Proof of proposition 6. By (11), when i is not a focus of π we have $\mu_i = 0$, If G is focal then μ is the singleton distribution on a. For G bifocal, rewrite (11) as

$$\mu_i = \pi_i \left(1 + \sum_j p_{ij}(H(j, \pi) - H(i, \pi)) \right).$$

Let $V_a = \{k \mid d(k, a) < d(k, b)$ When $i \in V_a$ is a neighbor of a, corollary 10 shows $H(i, \pi) - H(a, \pi) = H(i, a)$. Equation (4) gives

$$\sum_{V_a} p_{ai}(H(i, \pi) - H(a, \pi)) = \sum_{V_a} p_{ai} H(i, a) = \frac{1}{d(a)}(H(a, b) - d(a)).$$

Considering the final neighbor b, equations (10) and (13) give

$$
\begin{aligned}
H(b, \pi) - H(a, \pi) &= H(b', b) - H(\pi, b) - H(a', a) + H(\pi, a) \\
&= H(b', a) + H(a, b) - H(a', b) - H(b, a) + H(\pi, a) - H(\pi, b) \\
&= H(b', a) - H(a', b)
\end{aligned}
$$

Thus our formula for μ_a becomes

$$
\begin{aligned}
\mu_a &= \frac{d(a)}{2|E|} \left[1 + \frac{1}{d(a)} \left(H(a, b) - d(a) + H(b', a) - H(a', b) \right) \right] \\
&= \frac{1}{2|E|} \left(H(a, b) + H(b', a) - H(a', b) \right) = \frac{1}{2|E|} \left(H(b', b) - H(a', b) \right).
\end{aligned}
$$

We can calculate μ_b directly as above, or use $\mu_b = 1 - \mu_a$ and equation (5). $\qquad\square$

Corollary 11 *For a focal tree, $T_{\text{forget}} = H(a', a)$. For a bifocal tree,*

$$
\begin{aligned}
T_{\text{forget}} &= H(a', \mu) = H(b', \mu) \\
&= \frac{1}{2|E|}(H(a, b)H(b, a) + H(a, b)H(a', b) + H(b, a)H(b', a)).
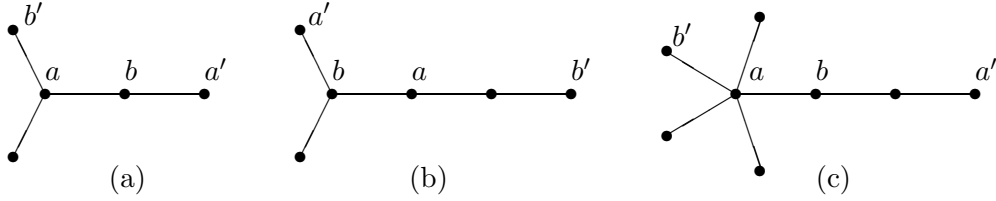\end{aligned}
$$

Figure 1: Trees with different foci achieving T_{bestmix}. (a) $H(a', b) = H(b', a)$ so $T_{\text{bestmix}} = H(a, \pi) = H(b, \pi)$. (b) $H(a', b) < H(b', a)$ so $T_{\text{bestmix}} = H(a, \pi)$. (c) $H(a', b) > H(b', a)$ so $T_{\text{bestmix}} = H(b, \pi)$.

Proof. Since a' and b' are mixing pessimal, $T_{\text{forget}} = H(a', \mu) = H(b', \mu)$ and the first statement is obvious. If G is bifocal, the following stopping rule is optimal from a' to μ: walk until you hit b, then stop with probability μ_b and walk to a with probability μ_a. Hence, $H(a', \mu) = H(a', b) + \mu_a H(b, a)$ and

$$T_{\text{forget}} = H(a', b) + \frac{H(b, a)}{2|E|}(H(b', b) - H(a', b)).$$

Equation (5) completes the proof. □

Proof of theorem 7. We quickly narrow our search down to the foci of the tree. Recall that a stopping rule is optimal if and only if it has a halting state. Lovász and Winkler [9] show that every mixing pessimal node is a halting state for an optimal rule from μ to π. Hence on a tree, both a' and b' are halting states for an optimal rule from μ to π. Therefore, for any node i, the rule "follow an optimal rule from i to the forget distribution μ and then follow an optimal rule from μ to π" has either a' or b' as a halting state. This rule is optimal and $H(i, \pi) = H(i, \mu) + H(\mu, \pi)$. We may minimize $H(i, \mu)$ rather than $H(i, \pi)$, which is clearly minimized by a focus of the tree.

If G has a unique focus, there is nothing to prove. Assume that G is bifocal with primary focus a and secondary focus b. Then

$$H(a, \mu) - H(b, \mu) = \mu_b H(a, b) - \mu_a H(b, a) = \frac{1}{2|E|}(H(a', b) - H(b', a))(H(a, b) + H(b, a)).$$

Thus $H(a, \mu) \geq H(b, \mu)$ if and only if $H(a', b) \geq H(b', a)$. □

The broom graphs $B_{2,2}$, $B_{2,3}$ and $B_{4,4}$ in figure 1 show that all three possibilities do occur.

5   Start-independent times

Start independent stopping rules also identify central nodes: we now prove proposition 8 and theorem 9, which show that the target distributions achieving $T_{\text{si}}(\pi)$ and T_{si} are concentrated on a barycenter and the foci of the tree, respectively. We continue to employ the notation $\sigma(S) = \sum_{k \in S} \sigma(k)$ and $\bar{S} = V \backslash S$ for $S \subset V$. The following lemma allows us to restrict ourselves to singleton targets.

Lemma 12 *Let σ and τ be distributions on the tree G. If τ has only one focus, denote this node by u. Otherwise let the foci u, v of τ satisfy $\sigma(V_u)\pi(V_u) \geq \sigma(V_v)\pi(V_v)$ where $V_u = \{i \in V \mid d(i, u) < d(i, v)\}$ and $V_v = \{i \in V \mid d(i, v) < d(i, u)\} = \bar{V}_u$. Then*

$$\sum_k \sigma_k H(k, \tau) \geq \sum_k \sigma_k H(k, u) = H(\sigma, u).$$

Proof.   If u is the only focus for τ then $\sum_k \sigma_k H(k, \tau) = \sum_k \sigma_k(H(k, u) + H(u, \tau)) \geq \sum_k \sigma_k H(k, u)$. If τ has two foci with $\sigma(V_u)\pi(V_u) \geq \sigma(V_v)\pi(V_v)$, equation (4) implies that $\sigma(V_u)H(u, v) \geq \sigma(V_v)H(v, u)$. By corollary 2,

$$
\begin{aligned}
\sum_V \sigma_k H(k, \tau) &= \sum_{V_u} \sigma_k(H(k, u) + H(u, \tau)) + \sum_{V_v} \sigma_k(H(k, v) + H(v, \tau)) \\
&= \sum_{V_u} \sigma_k H(k, u) + \sum_{V_v} \sigma_k H(k, v) + \sigma(V_u)H(u, \tau) + \sigma(V_v)H(v, \tau) \\
&= \sum_V \sigma_k H(k, u) - \sigma(V_v)H(v, u) + \sigma(V_u)H(u, \tau) + \sigma(V_v)H(v, \tau).
\end{aligned}
$$

For any rule from u to τ, we must step from u to V_v with probability $\tau(V_v)$ before halting, hence $H(u, \tau) \geq \tau(V_v)H(u, v)$. Likewise, $H(v, \tau) \geq \tau(V_u)H(v, u)$. Therefore

$$
\begin{aligned}
\sigma(V_u)H(u, \tau) + \sigma(V_v)H(v, \tau) &\geq \sigma(V_u)\tau(V_v)H(u, v) + \sigma(V_v)\tau(V_u)H(v, u) \\
&\geq \sigma(V_v)H(v, u)
\end{aligned}
$$

so $\sum_V \sigma_k H(k, \tau) \geq \sum_V \sigma_k H(k, u)$.   □

Proof of proposition 8. Taking $\sigma = \pi$ in lemma 12, a singleton target achieves $T_{\text{si}}(\pi)$. Proposition 1 shows that $T_{\text{si}}(\pi) = \min_i \sum_k \pi_k H(k, i) = \min_i H(\pi, i) = H(\pi, c)$.   □

Obviously $T_{\text{si}}(\pi) \leq T_{\text{reset}}$ and in fact $T_{\text{si}}(\pi)$ can be arbitrarily small in comparison. Consider the tree consisting of a path of length $2k$ with k^4 leaves connected to the central node c. The focus, center and barycenter of G are all located at c, so the forget distribution is concentrated on this central node. $T_{\text{forget}} = k^2$ while $H(\pi, c)$ becomes arbitrarily close to $1/2$ for large k.

On the other hand, theorem 9 states that the forget time and the start-independent time of a tree are identical. The theorem is clearly true for the path on two nodes, so we restrict our proof to trees on three or more nodes. We make the important observation that T_{si} need not be achieved by a unique pair of distributions. For example, consider a star graph with n leaves. Clearly $T_{\text{si}} \leq 1$ since we may always choose the central node c as our target. For any distribution σ concentrated on the leaf set such that $\sigma_i \leq (2n-1)/2n$ for every node i, we have $T_{\text{si}} = \min_j H(\sigma, j) = H(\sigma, c) = 1$.

We prove theorem 9 by constructing a particular initial distribution ϕ concentrated on two leaves such that $\min_j H(\phi, j) = T_{\text{si}}$. Once we have identified such a ϕ, we show that we may choose the target node to be a focus of G.

Lemma 13 *The node u is such that $H(\sigma, u) = \min_j H(\sigma, j)$ if and only if for each neighbor v_i of u, $1 \leq i \leq d(u)$, we have*

$$\sigma(V_i) H(v_i, u) \leq \sigma(\bar{V}_i) H(u, v_i)$$

where $V_i = \{k \in V \mid d(k, v) < d(k, u)\}$. We have equality if and only if $H(\sigma, u) = H(\sigma, v_i)$ so that v_i is also a best target for σ.

Proof. For any neighbor v_i of u,

$$H(\sigma, v_i) = \sum_{\bar{V}_i} \sigma_k H(k, u) + \sigma(\bar{V}_i) H(u, v_i) + \sum_{V_i} \sigma_k H(k, v_i).$$

We have $\sigma(V_i) H(v_i, u) \leq \sigma(\bar{V}_i) H(u, v_i)$ if and only if $H(\sigma, v_i) \geq \sum_{\bar{V}_i} \sigma_k H(k, u) + \sigma(V_i) H(v_i, u) + \sum_{V_i} \sigma_k H(k, v_i) = H(\sigma, u)$. Furthermore, equality holds in the first if and only if equality holds in the second. $\square$

We employ the following terminology for the remainder of the section. Let ϕ be a distribution, $S_\phi = \{v \mid \phi_v > 0\} \subset V$ and u a node such that $H(\phi, u) = \min_j H(\phi, j) = T_{\text{si}}$. Let $v_1, v_2, \ldots, v_{d(u)}$ be the neighbors of u. Let $G_i = (V_i, E_i)$ be the connected component of $G \backslash \{u\}$ with $v_i \in G_i$, and let $w_i \in V_i$ be a leaf such that $H(w_i, u) = \max_{j \in V_i} H(j, u)$.

Proof of theorem 9. Let $T_{\text{si}} = \sum_i \phi_i H(i, u)$.

Claim 1: S_ϕ intersects more than one component of $G \backslash \{u\}$.

Assume that S_ϕ intersects exactly one of $V_1, V_2, \ldots V_{d(u)}$.

CASE 1: $\phi_u = 0$. We may assume $S_\phi \subset G_1$. Then $H(\phi, v_1) < H(\phi, u)$ where $v_1 \in V_1$ is adjacent to u, contradicting $\min_j H(\phi, j) = H(\phi, u)$.

Case 2: $\phi_u \neq 0$. We may assume $S_\phi \subset V_1 \cup \{u\}$. We have $\phi(\bar{V}_1)H(u, v_1) \geq \phi(V_1)H(v_1, u)$ by lemma 13. If we have equality here, we may take v_1 as our target, proving the lemma. If we have strict inequality, there exists $\epsilon > 0$ such that the distribution ϕ' defined by $\phi'_u = \phi_u - \epsilon$, $\phi'_{v_1} = \phi_{v_1} + \epsilon$ and $\phi'_i = \phi_i$ otherwise satisfies $\phi'(\bar{V}_1)H(u, v_1) > \phi'(V_1)H(v_1, u)$. Lemma 13 shows that $\min_i H(\phi', i) = H(\phi', u)$, while $H(\phi', u) > H(\phi, u) = T_{\text{si}}$, a contradiction.

Claim 2: ϕ may be chosen so that S_ϕ is contained in the leaf set $\{w_1, w_2, \ldots, w_{d(u)}\}$.

Assume instead that $S_\phi \not\subset \{w_1, w_2, \ldots, w_{d(u)}\}$.

Case 1: $\phi_u = 0$. Let ϕ' be the distribution given by $\phi'_{w_i} = \phi(V_i)$ for $1 \leq i \leq d(u)$ and zero elsewhere. Lemma 13 and $H(\phi, u) = \min_i H(\phi, i)$ imply that $H(\phi', u) = \min_i H(\phi', i)$ as well. Clearly, $H(\phi', u) \geq H(\phi, u)$, so we may use ϕ' in place of ϕ.

Case 2: $\phi_u \neq 0$. Arguing similarly to case 1, we may choose ϕ so that $S_\phi \subset \{u, w_1, w_2, \ldots, w_{d(u)}\}$. If $\phi(\bar{V}_i)H(u, v_i) = \phi(V_i)H(v_i, u)$ for some i, we may take v_i as our target node in lieu of u by lemma 13. In this case, $\phi_{v_i} = 0$ and we have reduced ourselves to case 1.

If $\phi(\bar{V}_i)H(u, v_i) > \phi(V_i)H(v_i, u)$ for $1 \leq i \leq d(u)$ then there exists $\epsilon > 0$ such that the distribution ϕ' given by $\phi'_u = \phi_u - \epsilon$, $\phi'_{w_1} = \phi_{w_1} + \epsilon$ and $\phi'_i = \phi_i$ otherwise satisfies $\phi'(\bar{V}_i)H(u, v_i) > \phi'(V_i)H(v_i, u)$ for $1 \leq i \leq d(u)$. By lemma 13, $\min_i H(\phi', i) = H(\phi', u) > H(\phi, u)$, a contradiction.

Claim 3: ϕ may be chosen to be concentrated on two leaves in $\{w_1, w_2, \ldots w_{d(u)}\}$.

Case 1: $\phi(\bar{V}_i)H(u, v_i) = \phi(V_i)H(v_i, u)$ for some i. By lemma 13, $H(\phi, u) = H(\phi, v_i)$. Notice that ϕ is supported in two components of $G \backslash \{v_i\}$, so using the proof of claim 2, we may define a new distribution ϕ' concentrated on two leaves such that $\min_k H(\phi', k) = H(\phi', v_i) \geq H(\phi, v_i) = H(\phi, u) = T_{\text{si}}$.

Case 2: $\phi(\bar{V}_i)H(u, v_i) > \phi(V_i)H(v_i, u)$ for all i. We show by induction that there exists a distribution ϕ' supported on two leaves such that $\min_i H(\phi', i) \geq H(\phi, u)$. The base case $|S_\phi| = 2$ is trivial. Assume that if $|S_\phi| = k - 1$ then there exists a ϕ' concentrated on two leaves satisfying $\min_j \sum_i \phi'_i H(i, j) = \sum_i \phi'_i H(i, u) = \sum_i \phi_i H(i, u) = T_{\text{si}}$.

Considering $|S_\phi| = k$, order $S_\phi = \{w_1, w_2, \ldots, w_k\}$ so that $w_i \in V_i$ and $H(w_1, u) \geq H(w_2, u) \geq \cdots \geq H(w_k, u)$. There exists $\epsilon > 0$ such that the distribution ϕ^* defined by $\phi^*_{w_1} = \phi_{w_1} + \epsilon$, $\phi^*_{w_k} = \phi_{w_k} - \epsilon$ and $\phi^*_i = \phi_i$ otherwise satisfies $\phi^*(\bar{V}_i)H(u, v_i) > \phi^*(V_i)H(v_i, u)$ for all i. If $H(w_1, u) > H(w_k, u)$ then by lemma 13, $\min_i H(\phi^*, i) = H(\phi^*, u) > H(\phi, u) = T_{\text{si}}$, a contradiction.

Otherwise, we have $H(w_i, u) = H(w_j, u)$ for all i, j. If there exists $0 < \epsilon < \phi_{w_k}$ such that $\phi^*(\bar{V}_k)H(u, v_k) = \phi^*(V_k)H(v_k, u)$ we have $H(\phi^*, v_k) = H(\phi^*, u) = H(\phi, u) = T_{\text{si}}$. Hence we may take ϕ^* as our starting distribution and v_k as our target node. The support of ϕ^* is contained in two connected components of $\bar{V}_k$, and so the proof of claim 2 shows that there exists a distribution ϕ' supported on two leaves such that $T_{\text{si}} = \min_i H(\phi', i)$. Finally, if $\phi^*(\bar{V}_k)H(u, v_k) > \phi^*(V_k)H(v_k, u)$ for all $0 \leq \epsilon \leq \phi_{w_k}$, taking $\epsilon = \phi_{w_k}$ we have a distribution supported on $k - 1$ leaves such that $\min_i H(\phi^*, i) = T_{\text{si}}$ and we are done by induction.

Claim 4: ϕ may be chosen so that S_ϕ concentrated on two leaves w_1 and w_2 such that $H(w_1, u) \geq H(w_2, u)$ and the target node u is a focus of G. If the tree G is focal, then ϕ is concentrated on two u-pessimal leaves. If the tree G is bifocal, then w_1 is u-pessimal, v_1 is the other focus of G and w_2 is v_1-pessimal. In this case, ϕ is given by $\phi_{w_1} = \pi(\bar{V}_1)$ and $\phi_{w_2} = \pi(V_1)$ and $H(\phi, v_1) = H(\phi, u) = T_{\text{si}}$.

By claim 3, we may assume without loss of generality that ϕ is concentrated on leaves $w_1 \in V_1$ and $w_2 \in V_2$ where $H(w_1, u) \geq H(w_2, u)$. In order to maximize the access time, the distribution ϕ must weight w_1 as much as possible while still keeping u as the best target node. By lemma 13, we must have $\phi_{w_1}H(v_1, u) \leq \phi_{w_2}H(u, v_1)$ and $\phi_{w_2}H(v_2, u) \leq \phi_{w_1}H(u, v_2)$. By (4),

$$\begin{cases} \phi_{w_1}\pi(V_1) \leq \phi_{w_2}\pi(\bar{V}_1) \\ \phi_{w_2}\pi(V_2) \leq \phi_{w_1}\pi(\bar{V}_2) \end{cases} \tag{15}$$

and the optimal choice is $\phi_{w_1} = \pi(\bar{V}_1)$ and $\phi_{w_2} = \pi(V_1)$. Note that

$$H(\phi, v_1) = H(\phi, u) \tag{16}$$

by lemma 13. By claim 2, $H(w_1, u) = \max_{V_1} H(i, u)$, so if $H(w_1, u) < H(u', u)$ then a u-pessimal node u' must lie in one of $V_2, V_3, \ldots V_{d(u)}$. Since $H(w_2, u) \leq H(w_1, u)$, w_2 cannot be u-pessimal, so consider the distribution ϕ' given by $\phi'_{w_1} = \phi_{w_1}$, $\phi'_{u'} = \phi_{w_2}$ and $\phi'_i = 0$ otherwise. This distribution ϕ' satisfies the inequalities analogous to (15) so $\min_i H(\phi', i) = H(\phi', u) > H(\phi, u) = T_{\text{si}}$, a contradiction. Hence, w_1 is u-pessimal. By a similar argument, w_2 must satisfy $H(w_2, u) = \max_{\bar{V}_1} H(i, u)$.

If u is the unique focus of G then w_2 must also be u-pessimal. If u is a focus of a bifocal G then v_1 must be the other focus of G and w_2 is v-pessimal. If u is not a focus, then the foci of G must be on the unique path between u and the u-pessimal node w_1. Hence w_1 is v_1-pessimal and $H(w_1, v_1) \geq H(w_2, v_1)$. By (16) we may take v_1 as our target node instead of u. If $H(w_1, v_1) = H(w_2, v_1)$ then v_1 must be the unique focus of G since w_1 is a v_1-pessimal node, and $T_{\text{si}} = H(\phi, v_1)$ as required.

Otherwise $H(w_1, v_1) > H(w_2, v_1)$. Let $H_1, H_2, \ldots H_{d(v)}$ be the connected components of $G\backslash\{v\}$ where $H_i = (W_i, F_i)$ and $w_1 \in W_1$ and $w_2 \in W_2$. Consider the distribution ϕ' given by $\phi'_{w_1} = \pi(\bar{W}_1)$, $\phi'_{w_2} = \pi(W_1)$ and $\phi'_i = 0$ otherwise. By lemma 13, $H(\phi', v_1) = \min_i H(\phi', i)$ and

$$
\begin{aligned}
H(\phi', v_1) &= \pi(\bar{W}_1)H(w_1, v_1) + \pi(W_1)H(w_2, v_1) \\
&= (\pi(\bar{V}_1) + \pi(\bar{W}_1))H(w_1, v_1) + \pi(W_1)H(w_2, v_1) \\
&> \pi(\bar{V}_1)H(w_1, v_1) + (\pi(\bar{W}_1) + \pi(W_1))H(w_2, v_1) \\
&= H(\phi, v_1) = H(\phi, u) = T_{\mathrm{si}},
\end{aligned}
$$

a contradiction, so u must be a focus of G.

Completion of proof. If G has a single focus a, then by claim 4, we may take ϕ to be concentrated on two a-pessimal leaves in different components of $G\backslash\{a\}$ and $T_{\mathrm{si}} = H(\phi, a) = H(a', a) = T_{\mathrm{forget}}$ by corollary 11. If G is bifocal with foci a and b, then by claim 4, we may take ϕ to be concentrated on an a-pessimal node a' and a b-pessimal node b'. Also, $T_{\mathrm{si}} = H(\phi, a) = H(\phi, b)$ by (16). Finally,

$$
\begin{aligned}
T_{\mathrm{si}} &= \mu_a T_{\mathrm{si}} + \mu_b T_{\mathrm{si}} = \mu_a H(\phi, a) + \mu_b H(\phi, b) \\
&= \phi_{a'}(\mu_a H(a', a) + \mu_b H(a', b)) + \phi_{b'}(\mu_a H(b', a) + \mu_b H(b', b)) \\
&= \phi_{a'} H(a', \mu) + \phi_{b'} H(b', \mu) = (\phi_{a'} + \phi_{b'})T_{\mathrm{forget}} = T_{\mathrm{forget}}
\end{aligned}
$$

by proposition 6. $\qquad\square$

References

[1] D. J. Aldous, Some inequalities for reversible Markov chains, *J. London Math. Soc.* **25** (1982), pp. 564–576.

[2] D. Aldous, L. Lovász and P. Winkler, Mixing times for uniformly ergodic Markov chains, *Stochastic Processes and their Applications* **71** (1997), pp. 165–185.

[3] A. Beveridge and L. Lovász, Random walks and the regeneration time, *J. Graph Theory* **29** (1998), pp. 57-62.

[4] G. Brightwell and P. Winkler, Extremal Cover Times for Random Walks on Trees, *J. Graph Theory* **14** (1990), pp. 547–554.

[5] D. Coppersmith, P. Tetali, and P. Winkler, Collisions among Random Walks on a Graph, *SIAM J. on Discrete Mathematics* **6** No. 3 (1993), pp. 363–374.

[6] P. Diaconis and J. Fill, Strong Stationary Times Via a New Form of Duality, *The Annals of Probability* **18** No. 4 (1990), pp. 1483–1522.

[7] L. Lovász, Random Walks on Graphs: a Survey, *Combinatorics, Paul Erdős is Eighty, Vol. II*, (eds. D. Miklós, V. T. Sós and T. Szőnyi) J. Bolyai Math. Soc. (1996), pp. 353–397.

[8] L. Lovász and P. Winkler, Efficient stopping rules for Markov chains, *Proceedings of the 27th ACM Symposium on the Theory of Computing* (1995), pp. 76–82.

[9] L. Lovász and P. Winkler, Reversal of Markov Chains and the Forget Time, *Cominatorics, Probability and Computing* **7** (1998), pp. 189–204.

[10] J. W. Moon, Random Walks on Random Trees, *J. Austral. Math. Soc.* **15** (1973), pp. 42–53.

[11] L. H. Pearce, Random Walks on Trees, *Discrete Math.* **30** (1980), pp. 269–276.

[12] J. W. Pitman, Occupation measures for Markov chains, *Adv. Appl. Prob.* **9** (1977), pp. 69-86.

[13] P. Tetali, Random Walks and Effective Resistance of Networks, *J. Theoretical Prob.* No. 1 (1991), pp. 101–109.

[14] P. Tetali and P. Winkler, Simultaneous Reversible Markov Chains, *Combinatorics, Paul Erdős is Eighty, Vol. I*, (eds. D. Miklós, V. T. Sós and T. Szőnyi) J. Bolyai Math. Soc. (1993), pp. 433–451.