An $O(n^2)$ bound for the relaxation time of a Markov chain on cladograms

by Jason Schweinsberg

Technical Report No. 572

Department of Statistics
University of California
367 Evans Hall # 3860
Berkeley, CA 94720-3860

March 1, 2000

Abstract

A cladogram is an unrooted tree with labeled leaves and unlabeled internal branchpoints of degree 3. Aldous has studied a Markov chain on the set of n-leaf cladograms in which each transition consists of removing a random leaf and its incident edge from the tree and then reattaching the leaf to a random edge of the remaining tree. Using coupling methods, Aldous has shown that a mixing-time parameter for this chain known as the relaxation time is $O(n^3)$. Here, we use a Poincaré inequality to prove an $O(n^2)$ bound for the relaxation time, establishing a conjecture of Aldous.

Keywords: mixing times, Markov chains, cladograms, continuum random tree, Poincaré inequalities.

1 Introduction

An *n-leaf cladogram*, as defined in [2], is an unrooted tree with n leaves and internal branchpoints of degree 3. The leaves are labeled, usually by the integers $1, \dots, n$, but the internal branchpoints are unlabeled. We denote the set of n-leaf cladograms by T_n. Each n-leaf cladogram contains $2n-3$ edges. Also, as noted in [2], the cardinality of T_n, which we denote by c_n, is given by

$$c_n = (2n-5)(2n-7)\dots(3)(1) = \frac{(2n)!}{(2n-1)(2n-3)2^n n!}. \tag{1}$$

Figure 1 below shows an example of a 7-leaf cladogram.

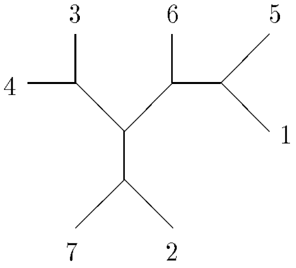

Figure 1: a 7-leaf cladogram

As mentioned in the introduction to [2], n-leaf cladograms are used in biology to represent the evolutionary relationship among n species. Since Markov chain Monte Carlo methods have been used to reconstruct cladograms from DNA data, it is of interest to study Markov chains on T_n. Here we will focus on a simple chain that was introduced in [2]. We first define the following two operations on cladograms:

(1) To *remove a leaf* i means to remove the edge from the leaf i to its branchpoint b, and then remove the branchpoint b to collapse two edges into one (see Figure 2).

(2) To *add a leaf* i *to the edge* g means to create a new branchpoint b in the interior of g and then add a new edge connecting b to the leaf i (see Figure 3).

Figure 2: removing the leaf i

Figure 3: adding the leaf i to the edge g

We now define a Markov chain $(X_t)_{t=0}^{\infty}$ with state space T_n. To define this chain, we construct X_{t+1} from X_t by removing a leaf chosen uniformly at random from X_t and then adding this leaf to an edge chosen uniformly at random from the remaining tree. For trees x and y in T_n, we write $y \sim x$ if $x \neq y$ and y can be obtained from x by removing one leaf of x and then adding it to an edge of the remaining tree. As observed in [2], the transition probabilities for this chain are given by

$$\begin{aligned} P(X_{t+1} = y | X_t = x) &= 1/n(2n-5) && \text{if } y \sim x \\ &= 1/(2n-5) && \text{if } y = x \end{aligned} \tag{2}$$

because there are n leaves that can be removed and $2n-5$ edges to which the leaf can be reattached. It follows that this chain has a symmetric transition matrix and therefore is reversible with a uniform stationary distribution.

As stated in [2], the chain $(X_t)_{t=0}^{\infty}$ is also irreducible and aperiodic. Therefore, the distribution of X_t converges to the uniform stationary distribution as $t \to \infty$. For designing Markov chain Monte Carlo algorithms, it is useful to know how long it takes for the chain to converge to the stationary distribution. There are different ways of formalizing this idea of "mixing time." See [7] for a detailed comparision of several mixing-time parameters. In this paper, we will consider the *relaxation time* τ_n for the Markov chain on T_n described above. The relaxation time is defined to be the reciprocal of the spectral gap, which means $\tau_n = 1/(1 - \lambda_{n,2})$, where $\lambda_{n,2}$ is the second-largest eigenvalue of the transition matrix for the chain.

A general discussion of the relaxation time for a reversible Markov chain and methods for bounding it is given in chapter 4 of [3]. In [2], Aldous used a coupling argument to show that for the Markov chain on n-leaf cladograms defined above, we have $\tau_n = O(n^3)$. He also conjectured that $\tau_n = O(n^2)$. In lectures given in a graduate course at U.C. Berkeley in the fall of 1999, Aldous outlined an alternative approach to bounding τ_n using a "distinguished paths" method. This approach did not yield a bound of $O(n^2)$, but we show in this paper that a variation of the "distinguished paths" method does give an $O(n^2)$ bound for τ_n. We state this result below.

Theorem 1 *Let τ_n be the relaxation time of the Markov chain $(X_t)_{t=0}^{\infty}$ on n-leaf cladograms defined above. Then $\tau_n = O(n^2)$.*

Theorem 1.1 of [2] shows that $\tau_n = \Omega(n^2)$, so it follows from Theorem 1 that $\tau_n = \Theta(n^2)$.

In section 2, we introduce the modification of the "distinguished paths" method that we will use to prove Theorem 1. We then prove Theorem 1 in section 3.

2 A modified distinguished paths method

A well-known method for upper-bounding the relaxation time of a reversible Markov chain is the method of distinguished paths, which is based on Poincaré inequalities. Diaconis and Stroock introduced this method in [4] and applied it to several examples. The method was further developed in [9] and [6], and the bounds were compared to bounds based on Cheeger's inequality in [5]. In this section, we present a variation that is useful for studying the Markov chain defined in the introduction.

Let $(Z_t)_{t=0}^{\infty}$ be an irreducible, aperiodic, reversible Markov chain with finite state space V and stationary distribution π. Define the transition probabilities by $p_{xy} = P(Z_{t+1} = y | Z_t = x)$. Let G be a graph with vertex set V and edge set E such that G has a directed edge from x to y if and only if $p_{xy} > 0$. For an edge e from x to y, define $Q(e) = \pi_x p_{xy}$. Let τ denote the relaxation time of the chain $(Z_t)_{t=0}^{\infty}$.

The following theorem follows from the proof of Theorem 2 in [6] and the first two sentences of the proof of Theorem 6 in [6], which show that the arguments used to prove Theorem 2 still hold when the paths between vertices are random.

Theorem 2 *Suppose, for all x and y in V, that γ_{xy} is a path in G, possibly random, from x to y such that no edge is traversed more than once. Let l be any function from E to $(0, \infty)$. Then,*

$$\tau \leq \max_{e \in E} \frac{1}{l(e)Q(e)} \sum_{x \in V} \sum_{y \in V} \pi_x \pi_y E[|\gamma_{xy}|_l] P(e \in \gamma_{xy}),$$

where the path length $|\gamma_{xy}|_l$ is defined by $|\gamma_{xy}|_l = \sum_{e \in \gamma_{xy}} l(e)$. In particular, if $E[|\gamma_{xy}|_l] \leq L$ for all $x, y \in V$, then

$$\tau \leq L \max_{e \in E} \frac{1}{l(e)Q(e)} \sum_{x \in V} \sum_{y \in V} \pi_x \pi_y P(e \in \gamma_{xy}).$$

To analyze the Markov chain defined in the introduction, we will need the following corollary of Theorem 2, which can be useful if one can only easily describe a short path γ_{xy} when y is in some subset B of V. We say that a length function from E to $(0, \infty)$ is *symmetric* if the length of a directed edge does not depend on the direction in which the edge is traversed.

Corollary 3 *Let B be a subset of V, and let $l : E \to (0, \infty)$ be a symmetric length function. Suppose, for all $x \in V$ and $y \in B$, that γ_{xy} is a path in G, possibly random, from x to y such that $E[|\gamma_{xy}|_l] \leq L$. Then,*

$$\tau \leq \frac{4L}{\pi(B)} \max_{e \in E} \frac{1}{l(e)Q(e)} \sum_{x \in V} \sum_{y \in B} \pi_x \pi_y P(e \in \gamma_{xy}).$$

Proof. For all x and z in V, define a path $\tilde{\gamma}_{xz}$ by choosing a random vertex Y in B such that $P(Y = y) = \pi_y / \pi(B)$ and then concatenating the paths γ_{xY} and γ_{zY}^r, where γ_{zY}^r is the path from z to Y traversed in the opposite direction. Denote by e^r the directed edge connecting the same two vertices as e but in the reverse direction. Then, using Theorem 1 for the first inequality, we have

$$\begin{aligned}
\tau &\leq \max_{e\in E} \frac{1}{l(e)Q(e)} \sum_{x\in V} \sum_{z\in V} \pi_x \pi_z E[|\tilde{\gamma}_{xz}|_l] P(e \in \tilde{\gamma}_{xz}) \\
&\leq \max_{e\in E} \frac{1}{l(e)Q(e)} \sum_{x\in V} \sum_{z\in V} \pi_x \pi_z E[|\gamma_{xY}|_l + |\gamma_{zY}|_l](P(e \in \gamma_{xY}) + P(e^r \in \gamma_{zY})) \\
&\leq 2L \max_{e\in E} \frac{1}{l(e)Q(e)} \sum_{x\in V} \sum_{z\in V} \pi_x \pi_z (P(e \in \gamma_{xY}) + P(e^r \in \gamma_{zY})) \\
&= 2L \max_{e\in E} \frac{1}{l(e)Q(e)} \left\{ \sum_{x\in V} \pi_x P(e \in \gamma_{xY}) \Big(\sum_{z\in V} \pi_z \Big) + \sum_{z\in V} \pi_z P(e^r \in \gamma_{zY}) \Big(\sum_{x\in V} \pi_x \Big) \right\} \\
&\leq 2L \max_{e\in E} \frac{1}{l(e)Q(e)} \sum_{x\in V} \pi_x P(e \in \gamma_{xY}) + 2L \max_{e\in E} \frac{1}{l(e)Q(e)} \sum_{z\in V} \pi_z P(e^r \in \gamma_{zY}).
\end{aligned}$$

For all $e \in E$, we have $l(e) = l(e^r)$ since l is symmetric and $Q(e) = Q(e^r)$ since the chain is reversible. Therefore,

$$\begin{aligned}
\tau &\leq 4L \max_{e\in E} \frac{1}{l(e)Q(e)} \sum_{x\in V} \pi_x P(e \in \gamma_{xY}) \\
&= 4L \max_{e\in E} \frac{1}{l(e)Q(e)} \sum_{x\in V} \pi_x \Big(\sum_{y\in B} \frac{\pi_y}{\pi(B)} P(e \in \gamma_{xy}) \Big) \\
&= \frac{4L}{\pi(B)} \max_{e\in E} \frac{1}{l(e)Q(e)} \sum_{x\in V} \sum_{y\in B} \pi_x \pi_y P(e \in \gamma_{xy}), \qquad (3)
\end{aligned}$$

as claimed. ■

3 Proof of the $O(n^2)$ bound

In this section, we prove Theorem 1. We first establish Lemmas 4 and 5, which give general results about uniform random cladograms. The natural inductive proof of formula (1) for the cardinality of T_n establishes that a uniform random cladogram can be built by inductively adding leaves to uniform random edges. Lemma 4 below emphasizes that this construction works for any ordering of leaves.

Lemma 4 *Let σ be a permutation of $\{1,\ldots,n\}$. Let x_3 be the unique 3-leaf cladogram with leaves labeled $\sigma(1)$, $\sigma(2)$, and $\sigma(3)$. For $4 \leq k \leq n$, inductively define x_k by adding a leaf labeled $\sigma(k)$ to an edge chosen uniformly at random from the edges of x_{k-1}. Then x_n is a uniform random n-leaf cladogram with leaves labeled $\{1,\ldots,n\}$.*

Proof. We proceed by induction. Clearly x_3 is a uniform random 3-leaf cladogram with leaves labeled $\sigma(1)$, $\sigma(2)$, and $\sigma(3)$. Suppose, for some $k \geq 4$, that x_{k-1} is a uniform random $(k-1)$-leaf cladogram with leaves labeled $\sigma(1), \ldots, \sigma(k-1)$. Let x be any k-leaf cladogram with leaves labeled $\sigma(1), \ldots, \sigma(k)$. Let y be the $(k-1)$-leaf cladogram obtained by removing the leaf labeled $\sigma(k)$ from x. Then there is an edge e of y such that x is obtained when a leaf labeled $\sigma(k)$ is added to the edge e. Furthermore, one can check that if $x_{k-1} \neq y$ or if $x_{k-1} = y$ and x_k is obtained from x_{k-1} by attaching a leaf labeled $\sigma(k)$ to any edge other than e, then $x_k \neq x$. Therefore, using the induction hypothesis and the fact that x_{k-1} has $2k-5$ edges, we obtain $P(x_k = x) = 1/(2k-5)c_{k-1} = 1/c_k$. It follows that x_k is a uniform random k-leaf cladogram with leaves labeled $\sigma(1), \ldots, \sigma(k)$. Hence, x_n is a uniform random n-leaf cladogram with leaves labeled $1, \ldots, n$. ■

Given a subset U of $\{1, \ldots, n\}$ and a cladogram $x \in T_n$, we define the *U-spanning tree of x* to be the tree obtained by successively removing from x all leaves whose labels are not in U. Note that the tree x_k in the construction of Lemma 4 is the $\{\sigma(1), \ldots, \sigma(k)\}$-spanning tree of x_n. Therefore, it follows from Lemma 4 that if U is a k-element subset of $\{1, \ldots, n\}$, then the U-spanning tree of a uniform random n-leaf cladogram is a uniform random k-leaf cladogram.

If $x \in T_n$, then any two leaves in x are connected by a unique path. Let $d(x)$ denote the *diameter* of x, which is the maximum number of edges in the path between any two leaves of x. Lemma 5 below states that the diameter of a typical n-leaf cladogram is $O(n^{1/2})$.

Lemma 5 *There exists a constant $A_1 < \infty$, not depending on n, such that if π denotes the uniform distribution on T_n and $B = \{x \in T_n : d(x) \leq A_1 n^{1/2}\}$, then $\pi(B) \geq 1/2$.*

Proof. Let x_n be a uniform random n-leaf cladogram. Following [2], we define a tree y_n by assigning random lengths to the $2n-3$ edges of x_n such that the joint density of the edge lengths is given by

$$g(l_1, \ldots, l_{2n-3}) = se^{-s^2/2}, \text{ where } s = \sum_{i=1}^{2n-3} l_i. \tag{4}$$

Note that the edge lengths are exchangeable and are independent of the shape of the tree. In the proof of Lemma 21 of [1], Aldous constructs a tree whose shape and edge lengths have the same distribution as y_n from the times $C_1 < C_2 < \ldots$ of a nonhomogeneous Poisson process on $(0, \infty)$ with rate $r(t) = t$. It follows from this construction that the sum of the lengths of the edges of y_n has the same distribution as C_{n-1}. This fact and the exchangeability of the edge lengths imply that there exists $A_2 > 0$ such that the expected length of any edge of y_n is at least $A_2 n^{-1/2}$ for all n.

Let v and w be leaves of x_n such that the number of edges on the path from v to w equals the diameter $d(x_n)$. Then the expected length, conditional on x_n, of the path in y_n from v to

w is at least $A_2 n^{-1/2} d(x_n)$. Therefore, the expected length of the longest path between any two leaves of y_n is at least $A_2 n^{-1/2} E[d(x_n)]$.

Next, let $(\mathcal{T}_\infty, \mu)$ be the Brownian continuum random tree, as defined in section 4.3 of of [1]. We may assume that $\mathcal{T}_\infty$ is constructed from the continuous function $f(t) = 2B_t$ as described in Theorem 13 of [1], where $(B_t)_{0 \leq t \leq 1}$ is a normalized Brownian excursion. Let $\mathcal{T}_n$ be the subtree of $\mathcal{T}_\infty$ spanned by $n-1$ leaves chosen at random from $\mathcal{T}_\infty$ according to the probability measure μ. Regard the root of $\mathcal{T}_n$ as a leaf, so $\mathcal{T}_n$ becomes an unrooted n-leaf tree. By Theorem 3 and Corollary 22 of [1], $\mathcal{T}_n$ has the same distribution as y_n. By Theorem 13 of [1], the length of any path between two vertices of $\mathcal{T}_n$ is at most $2 \max_{t \in [0,1]} f(t) = 4M_1$, where $M_1 = \max_{t \in [0,1]} B_t$. Therefore, $E[d(x_n)] \leq 4A_2^{-1} E[M_1] n^{1/2}$. Since $E[M_1] < \infty$, as shown, for example, in [8], it follows from Markov's inequality that the conclusion of Lemma 5 holds with $A_1 = 8A_2^{-1} E[M_1]$. ■

Now, we work towards using Corollary 3 to obtain an $O(n^2)$ bound for the relaxation time τ_n of the Markov chain on cladograms defined in the introduction. We use the notation from section 2, so in this case $V = T_n$, the stationary distribution π is uniform on T_n, and G is the graph with vertex set V such that there is an edge in G from x to y if and only if $y \sim x$.

For all $x, y \in V$, we now construct a random path γ_{xy} in G from x to y. First, let σ be a uniform random permutation of $\{1, \ldots, n\}$. Let $x_1 = x$. Define x_2 to be the cladogram that we obtain by removing the leaf labeled $\sigma(2)$ from x_1 and then adding it to the edge incident to $\sigma(1)$. Let S_2 denote the rooted subtree of x_2 whose root is a branchpoint of x_2 and whose only leaves are $\sigma(1)$ and $\sigma(2)$. Let S_2' be the unrooted tree obtained by regarding the root of S_2 as a leaf and labeling it $\sigma(n)$. Note that S_2' is the $\{\sigma(1), \sigma(2), \sigma(n)\}$-spanning tree of y because there is only one 3-leaf cladogram. Figure 4 below gives an example.

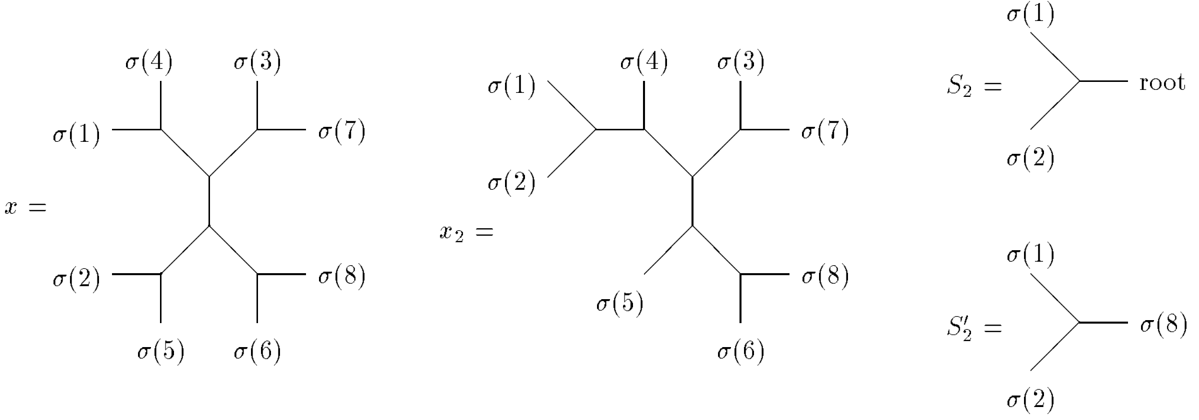

Figure 4: an example of x, x_2, S_2, and S_2'

Suppose, for some $k \in \{2, 3, \ldots, n-2\}$, that we have defined cladograms $x_2, \ldots, x_k$ and x_k has a rooted subtree S_k with leaves $\sigma(1), \ldots, \sigma(k)$ such that if we define S_k' by regarding the root of S_k as a leaf and labeling it $\sigma(n)$, then S'_k equals the $\{\sigma(1), \ldots, \sigma(k), \sigma(n)\}$-spanning tree of y. Then, we can define x_{k+1} by removing the leaf $\sigma(k+1)$ from x_k and adding it to the unique edge of S_k such that, after the leaf is added, the new subtree S_{k+1} has the property that when the root is labeled $\sigma(n)$ and regarded as a leaf, the new unrooted tree S'_{k+1} is the $\{\sigma(1), \ldots, \sigma(k+1), \sigma(n)\}$-spanning tree of y. See Figure 5 below, which shows the trees x_3 and x_4 defined when constructing the path from x to y, where x is the same tree as in Figure 4. Note that in going from x_3 to x_4, the leaf $\sigma(4)$ is removed and then added to the edge incident to $\sigma(2)$. Also, note that the $\{\sigma(1), \sigma(2), \sigma(3), \sigma(8)\}$-spanning tree of x_3 is the same as the $\{\sigma(1), \sigma(2), \sigma(3), \sigma(8)\}$-spanning tree of y and the $\{\sigma(1), \sigma(2), \sigma(3), \sigma(4), \sigma(8)\}$-spanning tree of x_4 is the same as the $\{\sigma(1), \sigma(2), \sigma(3), \sigma(4), \sigma(8)\}$-spanning tree of y.

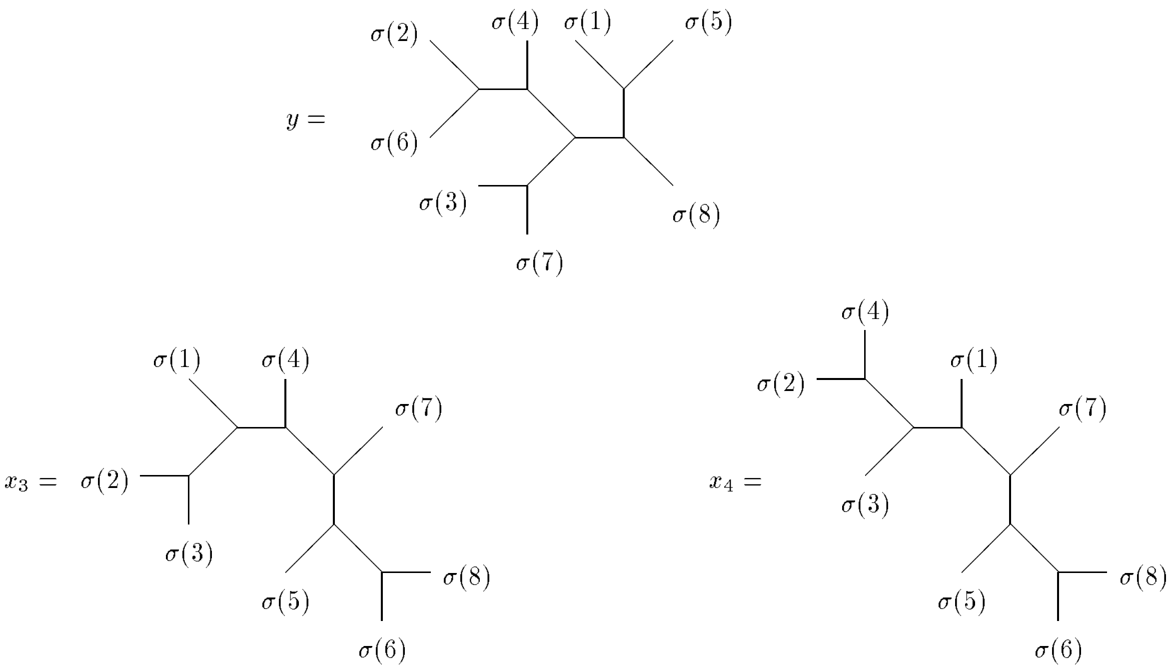

Figure 5: the trees x_3 and x_4

Note that when $x_1, \ldots, x_{n-1}$ are defined by this process, the tree S_{n-1} is just x_{n-1} with $\sigma(n)$ regarded as the root, so $x_{n-1} = S'_{n-1} = y$. Thus, we have inductively defined random cladograms $x_1, \ldots, x_{n-1}$ such that $x_1 = x$, $x_{n-1} = y$, and, for $1 \leq k \leq n-2$, either $x_k = x_{k+1}$ or $x_{k+1} \sim x_k$. We now define γ_{xy} to be the path consisting of all edges e that connect x_k to x_{k+1} for some k. If e is the edge from x_k to x_{k+1}, then we say $e \in \gamma_{xy}$ at step k.

Let $l(e) = 1$ for each edge $e \in E$. Then $|\gamma_{xy}|_l \leq n-2$ for all $x, y \in V$. Thus, Corollary 3 implies that

$$\tau_n \leq \frac{4(n-2)}{\pi(B)} \max_{e \in E} \frac{1}{Q(e)} \sum_{x \in V} \sum_{y \in B} \pi_x \pi_y P(e \in \gamma_{xy}),$$

where $B = \{x \in T_n : d(x) \leq A_1 n^{1/2}\}$ as defined in Lemma 5. By Lemma 5, $\pi(B) \geq 1/2$. Since π is uniform on V, we have $\pi_x = c_n^{-1}$ for all $x \in V$. By (2), we have $p_{xy} = 1/n(2n-5)$ whenever $y \sim x$. Thus, for all edges e in G, we have $Q(e) = \pi_x p_{xy} = 1/c_n n(2n-5)$. It follows that

$$\begin{aligned} \tau_n &\leq 8(n-2) \max_{e \in E} c_n n(2n-5) \sum_{x \in V} \sum_{y \in B} \pi_x \pi_y P(e \in \gamma_{xy}) \\ &\leq 16 n^3 c_n \max_{e \in E} \sum_{x \in V} \sum_{y \in B} \pi_x \pi_y P(e \in \gamma_{xy}). \qquad (5) \end{aligned}$$

For all $e \in E$, define $K(e) = \{k : P(e \in \gamma_{xy} \text{ at step } k) > 0 \text{ for some } x \in V, y \in B\}$. Now, fix an edge e. Let v and w be the cladograms such that e is the edge from v to w, and let i and g be the leaf and edge respectively in v such that w is obtained from v by removing the leaf i and then adding it to the edge g. Let p be the unique path in v that begins by traversing the edge incident to i and ends by traversing the edge g. If $e \in \gamma_{xy}$ at step k, then $v = x_k$ and $w = x_{k+1}$. Also, v has a rooted tree S_k, as defined above in the description of the construction of γ_{xy}, which contains k leaves and the edge g. Note that the root r of S_k must be one of the branchpoints on the path p (see Figure 6).

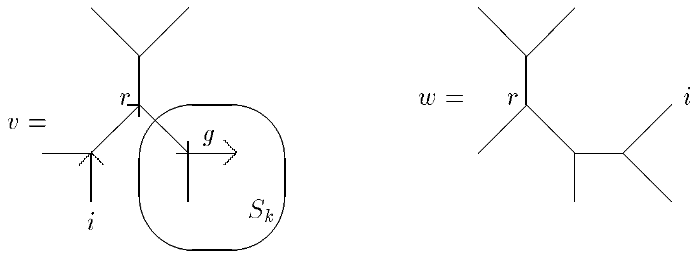

Figure 6: arrows mark the path p; the subtree S_k is circled.

If $y \in B$, then y has diameter at most $A_1 n^{1/2}$. Since S'_k is the $\{\sigma(1), \ldots, \sigma(k), \sigma(n)\}$-spanning tree of y and S'_k contains g, the portion of the path p that starts at r and ends after traversing g has length at most $A_1 n^{1/2}$. Also, there is at most one subtree rooted at a given branchpoint on the path p that could be S_k. Therefore, the number of integers in $K(e)$ is at most $A_1 n^{1/2}$. Thus,

$$\begin{aligned} \tau_n &\leq 16 n^3 c_n \max_{e \in E} \sum_{x \in V} \sum_{y \in B} \pi_x \pi_y \Bigg(\sum_{k \in K(e)} P(e \in \gamma_{xy} \text{ at step } k) \Bigg) \\ &\leq 16 A_1 n^{7/2} c_n \max_{e \in E} \max_{k \in K(e)} \sum_{x \in V} \sum_{y \in V} \pi_x \pi_y P(e \in \gamma_{xy} \text{ at step } k). \qquad (6) \end{aligned}$$

Note that

$$\sum_{x \in V} \sum_{y \in V} \pi_x \pi_y P(e \in \gamma_{xy} \text{ at step } k) \tag{7}$$

is the probability that $e \in \gamma_{xy}$ at step k when x and y are chosen uniformly at random from T_n. We now assume x and y are uniform random n-leaf cladograms and bound the expression in (7) for fixed $e \in E$ and $k \in K(e)$. Define v, w, i, p, and g as in the previous paragraph. We have seen that, since $k \in K(e)$, the tree v must have a rooted subtree S_k with k leaves which contains the edge g and has its root on the path p. Also, for any fixed k, v has only one such subtree. Now construct a random path γ_{xy} as described earlier. We have $e \in \gamma_{xy}$ at step k if and only if $x_k = v$ and $x_{k+1} = w$, where x_k and x_{k+1} are as defined in the construction of γ_{xy}. Let σ be the random permutation of $\{1, \ldots, n\}$ required for the construction of γ_{xy}. We claim that if $x_k = v$ and $x_{k+1} = w$, then the following three events must occur:

(a) The leaves of S_k are $\sigma(1), \ldots, \sigma(k)$, and $\sigma(k+1) = i$.
(b) The $\{\sigma(1), \sigma(k+1), \ldots, \sigma(n)\}$-spanning tree of x is the same as the $\{\sigma(1), \sigma(k+1), \ldots, \sigma(n)\}$-spanning tree of v.
(c) The $\{\sigma(1), \ldots, \sigma(k+1), \sigma(n)\}$-spanning tree of y is the same as the $\{\sigma(1), \ldots, \sigma(k+1), \sigma(n)\}$-spanning tree of w.

Event (a) must occur because, to obtain x_{k+1} from x_k, the leaf $\sigma(k+1)$ is removed and then added to an edge in a subtree of x_k whose k leaves are $\sigma(1), \ldots, \sigma(k)$. Event (b) must occur because none of the leaves $\sigma(1), \sigma(k+1), \ldots, \sigma(n)$ is moved in the first $k-1$ steps in the construction of γ_{xy}, so the $\{\sigma(1), \sigma(k+1), \ldots, \sigma(n)\}$-spanning tree of x is the $\{\sigma(1), \sigma(k+1), \ldots, \sigma(n)\}$-spanning tree of x_k. Also, event (c) must occur because the leaves $\sigma(1), \ldots, \sigma(k+1), \sigma(n)$ are not moved again after the kth step in the construction, so the $\{\sigma(1), \ldots, \sigma(k+1), \sigma(n)\}$-spanning tree of y is the same as the $\{\sigma(1), \sigma(k+1), \ldots, \sigma(n)\}$-spanning tree of x_{k+1}.

Since σ is a uniform random permutation of $\{1, \ldots, n\}$, we have that $\sigma(1), \ldots, \sigma(k)$ are the leaves of S_k with probability $1/\binom{n}{k}$, and conditional on the event that $\sigma(1), \ldots, \sigma(k)$ are the leaves of S_k, we have $\sigma(k+1) = i$ with probability $1/(n-k)$. By Lemma 4, since x is chosen uniformly at random from T_n, the $\{\sigma(1), \sigma(k+1), \ldots, \sigma(n)\}$-spanning tree of x is a uniform random element of T_{n-k+1}. Therefore, the probability of event (b) is $1/c_{n-k+1}$. Likewise, the $\{\sigma(1), \ldots, \sigma(k+1), \sigma(n)\}$-spanning tree of y is a uniform random element of T_{k+2}, so the probability of event (c) is $1/c_{k+2}$. Therefore, if $k \in K(e)$, then

$$\begin{aligned} \sum_{x \in V} \sum_{y \in V} \pi_x \pi_y P(e \in \gamma_{xy} \text{ at step } k) &\leq \frac{1}{\binom{n}{k}(n-k)c_{k+2}c_{n-k+1}} \\ &= \frac{1}{\binom{n}{k}(n-k)(2k-1)(2k-3)(2(n-k)-3)c_k c_{n-k}} \end{aligned} \tag{8}$$

because $c_{k+2} = (2k-1)(2k-3)c_k$ and $c_{n-k+1} = (2(n-k)-3)c_{n-k}$ by (1). Combining (6) and (8), we get

$$\tau_n \le 16A_1 n^{7/2} \max_{e\in E} \max_{k\in K(e)} \frac{c_n}{\binom{n}{k}(n-k)(2k-1)(2k-3)(2(n-k)-3)c_k c_{n-k}}. \tag{9}$$

By Stirling's formula, we have

$$\frac{(2n)!}{n!} \approx \frac{\sqrt{2\pi}(2n)^{2n+1/2}e^{-2n}}{\sqrt{2\pi}n^{n+1/2}e^{-n}} = 2^{2n+1/2}n^n e^{-n},$$

where $\approx$ means that the ratio of two sides is bounded away from zero and infinity as n varies. Therefore, using (1), we have

$$c_n = \frac{(2n)!}{(2n-1)(2n-3)2^n n!} \approx \frac{2^{2n+1/2}n^n e^{-n}}{(2n-1)(2n-3)2^n} \approx 2^{n-3/2}n^{n-2}e^{-n}.$$

Also, we have

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \approx \frac{\sqrt{2\pi}n^{n+1/2}e^{-n}}{(\sqrt{2\pi}k^{k+1/2}e^{-k})(\sqrt{2\pi}(n-k)^{n-k+1/2}e^{-(n-k)})} = \frac{n^{n+1/2}}{\sqrt{2\pi}k^{k+1/2}(n-k)^{n-k+1/2}},$$

where here $\approx$ means that the ratio of the two sides is bounded away from zero and infinity as n and k vary. Thus, there exists a constant $A_3 > 0$ such that for all $n \in \mathbb{N}$ and $k \in \{1, \ldots, n-1\}$, we have

$$\begin{aligned}
&\frac{c_n}{\binom{n}{k}(n-k)(2k-1)(2k-3)(2(n-k)-3)c_k c_{n-k}} \\
\le\ & \frac{A_3}{k^2(n-k)^2} \frac{k^{k+1/2}(n-k)^{n-k+1/2}}{n^{n+1/2}} \frac{2^{n-3/2}n^{n-2}e^{-n}}{(2^{k-3/2}k^{k-2}e^{-k})(2^{n-k-3/2}(n-k)^{n-k-2}e^{-(n-k)})} \\
=\ & \frac{2\sqrt{2}A_3}{k^2(n-k)^2}\left(\frac{k(n-k)}{n}\right)^{5/2} = 2\sqrt{2}A_3 n^{-5/2}(k(n-k))^{1/2} \le 2\sqrt{2}A_3 n^{-3/2}
\end{aligned} \tag{10}$$

for all e and all $k \in K(e)$. Thus, from (9) and (10), we have $\tau_n \le 32\sqrt{2}A_1A_3n^2$. Hence, $\tau_n = O(n^2)$, as claimed.

Acknowledgments

The author would like to thank David Aldous for suggesting this problem, giving a preliminary outline of the argument used to prove the main result, and commenting on earlier drafts of this paper.

References

[1] D. Aldous. The continuum random tree III. *Ann. Probab.*, 21:248–289, 1993.

[2] D. Aldous. Mixing time for a Markov chain on cladograms. To appear in *Combin. Probab. Comput.*, 2000.

[3] D. Aldous and J. Fill. Reversible Markov chains and random walks on graphs. Book in preparation, 1999.

[4] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1:36–61, 1991.

[5] J. Fulman and E. Wilmer. Comparing eigenvalue bounds for Markov chains: When does Poincaré beat Cheeger? *Ann. Appl. Probab.*, 9:1–13, 1999.

[6] N. Kahale. A semidefinite bound for the mixing rates of Markov chains. *Random Structures Algorithms*, 11:299–313, 1997.

[7] L. Lovász and P. Winkler. Mixing times. In D. Aldous and J. Propp, editors, *Microsurveys in Discrete Probability, DIMACS Series in Discrete Mathematics and Theoretical Computer Science.* Vol. 41, pp. 85-134. American Mathematical Society, 1998.

[8] J. Pitman and M. Yor. The law of the maximum of a Bessel bridge. *Electron. J. Probab.*, 4:1–35, 1999.

[9] A. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.*, 1:351–370, 1992.