



# *An $O(n^2)$ Bound for the Relaxation Time of a Markov Chain on Cladograms*

**Jason Schweinsberg**

*University of California, Berkeley, California; email: jasons@math.cornell.edu*

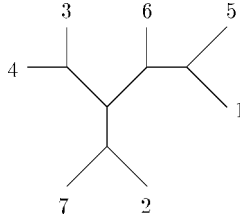
*Received 9 March 2000; revised 9 May 2001; accepted 24 July 2001*

**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 showed that the relaxation time (i.e., the inverse of the spectral gap) for this chain is  $O(n^3)$ . Here, we use a method based on distinguished paths to prove an  $O(n^2)$  bound for the relaxation time, establishing a conjecture of Aldous. © 2001 John Wiley & Sons, Inc. *Random Struct. Alg.*, 20, 59–70, 2001

*Key Words:* mixing times; cladograms; Markov chains; distinguished paths; Poincaré inequalities; continuum random tree

## 1. INTRODUCTION

An  $n$ -leaf cladogram 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. The cardinality of  $T_n$ , which we denote by



**Fig. 1.** A 7-leaf cladogram.

$c_n$ , is given by

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

Figure 1 shows an example of 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. Because Markov chain Monte Carlo methods have been used to reconstruct cladograms from DNA data, it is of interest to study Markov chains on  $T_n$ . Aldous [2] introduced a simple Markov chain on cladograms. He proved that the relaxation time (i.e., the inverse of the spectral gap) is  $O(n^3)$  and conjectured that it is  $O(n^2)$ . The purpose of this article is to prove Aldous's conjecture. Here we recall for reference Aldous's definitions and some basic properties of the chain.

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 the 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).

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  $y \neq x$  and  $y$  can be obtained from  $x$  by removing one leaf of  $x$  and then adding it to an edge of the remaining tree. Because there are  $n$  leaves that can be removed and  $2n - 5$  edges to which the leaf can be reattached, the transition probabilities for this chain are given by

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



**Fig. 2.** Removing the leaf  $i$ .



**Fig. 3.** Adding the leaf  $i$  to the edge  $g$ .

It follows that this chain has a symmetric transition matrix and therefore is reversible with a uniform stationary distribution.

The chain  $(X_t)_{t=0}^\infty$  is irreducible and aperiodic. Therefore, the distribution of  $X_t$  converges to the uniform stationary distribution as  $t \rightarrow \infty$ . For designing Markov chain Monte Carlo algorithms, it is useful to know the time it takes for the chain to converge to the stationary distribution. There are different ways of formalizing this idea of “mixing time.” See [11] for a detailed comparison of several mixing-time parameters. In this article, we will consider the *relaxation time*  $\tau_n$  for the Markov chain on  $T_n$  described above. The relaxation time is defined to be the inverse of the spectral gap. That is,  $\tau_n = 1/(1 - \lambda_{n,2})$ , where  $\lambda_{n,2}$  is the second-largest eigenvalue of the transition matrix for the chain.

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

**Theorem 1.** *Let  $\tau_n$  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 prove Theorem 1 in Section 3 and give some brief concluding remarks in Section 4.

## 2. A MODIFIED DISTINGUISHED PATHS METHOD

Let  $(Z_t)_{t=0}^\infty$  be an irreducible, aperiodic, reversible Markov chain with finite state space  $V$  and stationary distribution  $\pi$ . 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  $p_{xy} > 0$ . Let  $Q(x, y) = \pi_x p_{xy}$ , and if  $e$  is an edge from  $x$  to  $y$ , let  $Q(e) = Q(x, y)$ . Let  $\tau$  denote the relaxation time of the chain  $(Z_t)_{t=0}^\infty$ .

A well known method for upper-bounding  $\tau$  is the method of distinguished paths. The idea is to choose paths connecting each pair of vertices in the graph  $G$ . If it

is possible to avoid bottlenecks, in which many paths go through the same edges of the graph, then one can obtain a good bound on the relaxation time. Jerrum and Sinclair [9] used distinguished paths to prove rapid mixing for a Markov chain that can be used to approximate the permanent of a matrix. Diaconis and Stroock [5] used distinguished paths to prove “Poincaré inequalities” and showed how to use these inequalities to get bounds on the relaxation time in various examples. Sinclair [15] and Kahale [10] obtained related inequalities. Fulman and Wilmer [8] compared the bounds from Poincaré inequalities to bounds derived from Cheeger’s inequality.

In this section, we present a modification of the Poincaré inequalities which is useful for studying the Markov chain described in the introduction. We first recall the following theorem, which is proved in chapter 4 of [4] and can be viewed as a probabilistic version of inequalities in [5] and [15].

**Theorem 2.** *Suppose, for all  $x$  and  $y$  in  $V$ , that  $\gamma_{xy}$  is a path in  $G$ , possibly random, from  $x$  to  $y$  such that no edge is traversed more than once. Then,*

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

where  $|\gamma_{xy}|$  is the number of edges in the path  $\gamma_{xy}$ .

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 define a short path  $\gamma_{xy}$  when  $y$  is in some subset  $B$  of  $V$ .

**Corollary 3.** *Let  $B$  be a subset of  $V$ . Suppose, for all  $x \in V$  and  $y \in B$ , that  $\gamma_{xy}$  is a path in  $G$ , possibly random, from  $x$  to  $y$  which has at most  $L$  edges. Then,*

$$\tau \leq \frac{4L}{\pi(B)} \max_{e \in E} \frac{1}{Q(e)} \sum_{x \in V} \sum_{y \in B} \pi_x \pi_y P(e \in \gamma_{xy}). \quad (3)$$

*Proof.* For all  $x$  and  $z$  in  $V$ , we define a random path  $\tilde{\gamma}_{xz}$  from  $x$  to  $z$  as follows. First, we choose a random vertex  $Y$  in  $B$  such that  $P(Y = y) = \pi_y / \pi(B)$ . Next, we concatenate the paths  $\gamma_{xY}$  and  $\gamma'_{zY}$ , where  $\gamma'_{zY}$  is the path from  $z$  to  $Y$  traversed in the opposite direction. Finally, we erase all loops, so that no edge of  $\tilde{\gamma}_{xz}$  is traversed more than once. Note that the length of  $\tilde{\gamma}_{xz}$  is at most  $2L$ . By Theorem 2,

$$\begin{aligned} \tau &\leq \max_{e \in E} \frac{1}{Q(e)} \sum_{x \in V} \sum_{z \in V} \pi_x \pi_z E[|\tilde{\gamma}_{xz}| 1_{\{e \in \tilde{\gamma}_{xz}\}}] \\ &\leq 2L \max_{e \in E} \frac{1}{Q(e)} \sum_{x \in V} \sum_{z \in V} \pi_x \pi_z P(e \in \tilde{\gamma}_{xz}). \end{aligned} \quad (4)$$

Denote by  $e^r$  the directed edge connecting the same two vertices as  $e$  but in the reverse direction. Then, we have

$$\begin{aligned} \tau &\leq 2L \max_{e \in E} \frac{1}{Q(e)} \sum_{x \in V} \sum_{z \in V} \pi_x \pi_z (P(e \in \gamma_{xY}) + P(e^r \in \gamma_{zY})) \\ &= 2L \max_{e \in E} \frac{1}{Q(e)} \left\{ \sum_{x \in V} \pi_x P(e \in \gamma_{xY}) \left( \sum_{z \in V} \pi_z \right) + \sum_{z \in V} \pi_z P(e^r \in \gamma_{zY}) \left( \sum_{x \in V} \pi_x \right) \right\} \\ &\leq 2L \max_{e \in E} \frac{1}{Q(e)} \sum_{x \in V} \pi_x P(e \in \gamma_{xY}) + 2L \max_{e \in E} \frac{1}{Q(e)} \sum_{z \in V} \pi_z P(e^r \in \gamma_{zY}). \end{aligned}$$

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

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

as claimed. ■

### 3. PROOF OF THE $O(n^2)$ BOUND

In this section, we use Corollary 3 to prove Theorem 1. Our strategy is as follows. We will construct a path between each pair of cladograms. To get a path from  $x$  to  $y$ , we will remove and reattach the leaves one at a time in such a way that at each stage, all the leaves that have already been moved are in the same relative position as they are in the tree  $y$ . We will show that the diameter (i.e., maximum distance between two leaves) of most cladograms is  $O(n^{1/2})$ , and we will define the set  $B$  to consist of all cladograms whose diameter is at most  $A_1 n^{1/2}$ , where  $A_1$  is a constant. We will then show that the right-hand side of (3) is  $O(n^2)$ , using the fact that the paths end at a tree whose diameter is not too large to show that no “edge” between two cladograms appears in too many paths.

Given a subset  $U$  of  $\{1, \dots, n\}$  and a cladogram  $x \in T_n$ , define the  $U$ -tree derived from  $x$  to be the tree obtained by removing from  $x$  all leaves whose labels are not in  $U$ . Note that we obtain the same tree regardless of the order in which the leaves are removed.

**Lemma 4.** *If  $U$  is a  $k$ -element subset of  $\{1, \dots, n\}$  and  $x$  is a uniform random  $n$ -leaf cladogram, then the  $U$ -tree derived from  $x$  is a uniform random  $k$ -leaf cladogram.*

*Proof.* 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. In fact, as can be easily checked, this construction works for

any ordering of the leaves. More precisely, suppose  $\sigma$  is a permutation of  $\{1, \dots, n\}$ . Let  $x_3$  be the unique 3-leaf cladogram with leaves labeled  $\sigma(1)$ ,  $\sigma(2)$ , and  $\sigma(3)$ . For  $j = 4, \dots, n$ , define  $x_j$  by adding a leaf labeled  $\sigma(j)$  to an edge chosen uniformly at random from the edges of  $x_{j-1}$ . Then each  $x_j$  is a uniform  $j$ -leaf cladogram with leaves labeled  $\sigma(1), \dots, \sigma(j)$ .

Now, choose  $\sigma$  such that  $U = \{\sigma(1), \dots, \sigma(k)\}$ . Then  $x_k$  is the  $U$ -tree derived from  $x_n$ . Because  $x_k$  is a uniform random  $k$ -leaf cladogram and  $x_n$  is a uniform random  $n$ -leaf cladogram, we see that the  $U$ -tree derived from a uniform random  $n$ -leaf cladogram is a uniform random  $k$ -leaf cladogram, as claimed. ■

If  $x \in T_n$ , then any two leaves in  $x$  are connected by a unique path that does not intersect itself. Let  $d(x)$  denote the *diameter* of  $x$ , which is the maximum number of edges in the path between 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  $\pi$  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

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

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  $C_1 < C_2 < \dots$  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 lengths, 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 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 [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 reduced subtree of  $\mathcal{T}_\infty$  having  $n - 1$  leaves, which are chosen at random from  $\mathcal{T}_\infty$  according to the probability measure  $\mu$ . 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,  $E[d(x_n)] \leq 4A_2^{-1} E[M_1] n^{1/2}$ . Because  $E[M_1] < \infty$ , as shown, for example, in [13],

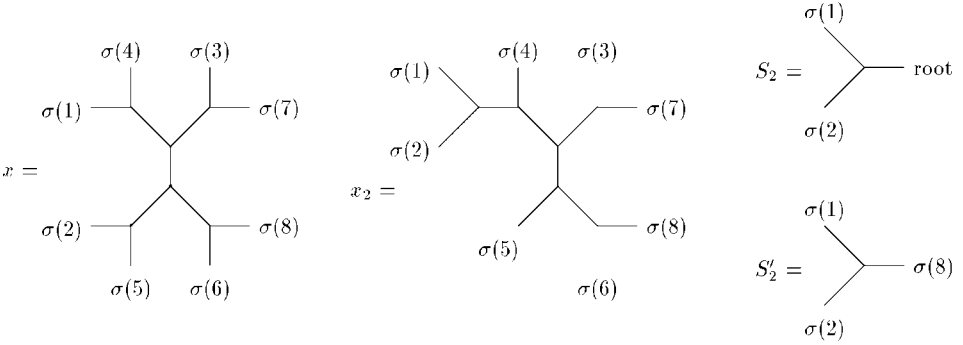


Fig. 4. An example of  $x$ ,  $x_2$ ,  $S_2$ , and  $S'_2$ .

Markov’s inequality implies that the condition of Lemma 5 holds with  $A_1 = 8A_2^{-1}E[M_1]$ . ■

We now work towards constructing our distinguished paths. We use the notation from Section 2, so  $V = T_n$  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  $y \sim x$ . The stationary distribution  $\pi$  is uniform on  $T_n$ .

For all  $x, y \in V$ , we will construct a random path  $\gamma_{xy}$  from  $x$  to  $y$ . First, let  $\sigma$  be a uniform random permutation of  $\{1, \dots, n\}$ . 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  $\{\sigma(1), \sigma(2), \sigma(n)\}$ -tree derived from  $y$  because there is only one 3-leaf cladogram. See Figure 4 for an example.

Next, suppose for some  $k \in \{2, 3, \dots, n - 2\}$ , we have defined cladograms  $x_2, \dots, x_k$ , and  $x_k$  has a rooted subtree  $S_k$  with leaves  $\sigma(1), \dots, \sigma(k)$  such that if we define  $S'_k$  by regarding the root of  $S_k$  as a leaf and labeling it  $\sigma(n)$ , then  $S'_k$  is the  $\{\sigma(1), \dots, \sigma(k), \sigma(n)\}$ -tree derived from  $y$ . 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  $\{\sigma(1), \dots, \sigma(k + 1), \sigma(n)\}$ -tree defined from  $y$ . See Figure 5, 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  $\{\sigma(1), \sigma(2), \sigma(3), \sigma(8)\}$ -tree derived from  $x_3$  is the same as the  $\{\sigma(1), \sigma(2), \sigma(3), \sigma(8)\}$ -tree derived from  $y$ , and the  $\{\sigma(1), \sigma(2), \sigma(3), \sigma(4), \sigma(8)\}$ -tree derived from  $x_4$  is the same as the  $\{\sigma(1), \sigma(2), \sigma(3), \sigma(4), \sigma(8)\}$ -tree derived from  $y$ . When  $x_1, \dots, x_{n-1}$  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  $x_1, \dots, x_{n-1}$  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  $\gamma_{xy}$  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  $k$ .

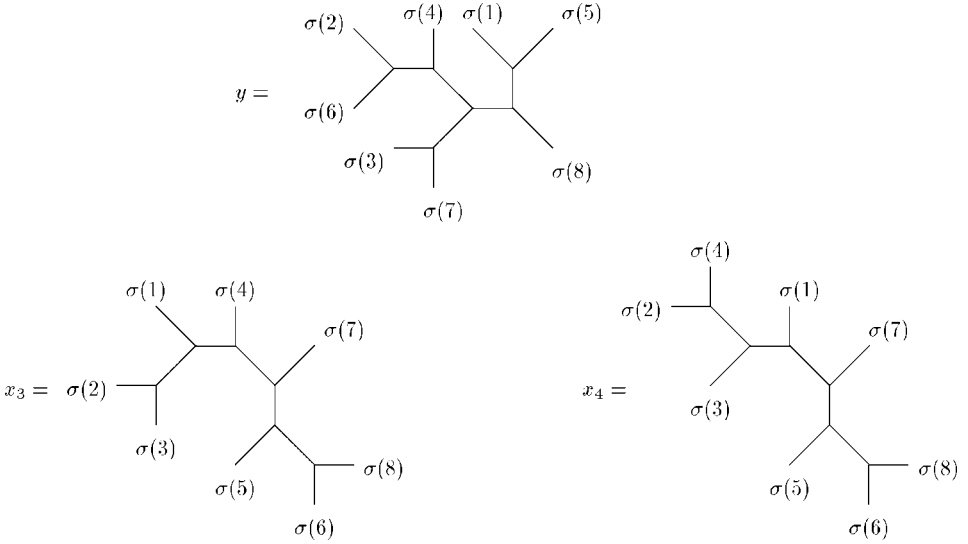


Fig. 5. The trees  $x_3$  and  $x_4$ .

Note that  $|\gamma_{xy}| \leq n - 2$  for all  $x, y \in V$ . Therefore, 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,  $\pi(B) \geq 1/2$ . Because  $\pi$  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  $e$  of  $G$ , we have  $Q(e) = \pi_x p_{xy} = 1/c_n n(2n - 5)$ . It follows that

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

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

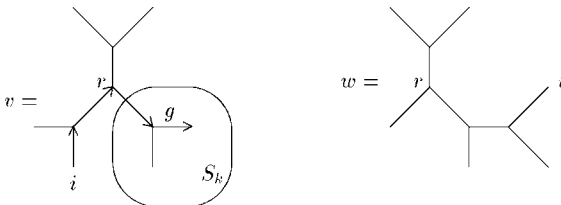


Fig. 6. Arrows mark the path  $p$ ; the subtree  $S_k$  is circled.

path in  $v$  that begins by traversing the edge incident to  $i$ , ends by traversing the edge  $g$ , and does not intersect itself. Suppose  $e \in \gamma_{xy}$  at step  $k$ . Then  $v = x_k$  and  $w = x_{k+1}$ . Also,  $v$  has a rooted subtree  $S_k$ , as defined above in the description of the construction of  $\gamma_{xy}$ , 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 Fig. 6).

If  $y \in B$ , then  $y$  has diameter at most  $A_1 n^{1/2}$ . Because  $S'_k$  is the  $\{\sigma(1), \dots, \sigma(k), \sigma(n)\}$ -tree derived from  $y$  and  $S'_k$  contains  $g$ , the portion of the path  $p$  that starts at  $r$  and ends after traversing  $g$  has length at most  $A_1 n^{1/2}$ . Therefore, there are at most  $A_1 n^{1/2}$  branchpoints on  $p$  that could be the root  $r$  of  $S_k$ . There are three subtrees of  $v$  rooted at each branchpoint, but only one of them contains the edge  $g$ . Therefore, there are at most  $A_1 n^{1/2}$  subtrees of  $v$  which could be the subtree  $S_k$ . Because  $S_k$  has  $k$  leaves, it follows that there are at most  $A_1 n^{1/2}$  possible values for  $k$ , so  $K(e)$  contains at most  $A_1 n^{1/2}$  integers. Thus,

$$\begin{aligned} \tau_n &\leq 16n^3 c_n \max_{e \in E} \sum_{x \in V} \sum_{y \in B} \pi_x \pi_y \left( \sum_{k \in K(e)} P(e \in \gamma_{xy} \text{ at step } k) \right) \\ &\leq 16A_1 n^{7/2} c_n \max_{e \in E} \max_{k \in K(e)} \sum_{x \in V} \sum_{y \in V} \pi_x \pi_y P(e \in \gamma_{xy} \text{ at step } k). \end{aligned} \tag{5}$$

Note that

$$\sum_{x \in V} \sum_{y \in V} \pi_x \pi_y P(e \in \gamma_{xy} \text{ at step } k) \tag{6}$$

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 independent uniform random  $n$ -leaf cladograms and bound the expression in (6) 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, because  $k \in K(e)$ , 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 let  $\sigma$  be a uniform random permutation of  $\{1, \dots, n\}$ , and construct a random path  $\gamma_{xy}$  from  $x$ ,  $y$ , and  $\sigma$  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  $\gamma_{xy}$ . 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  $\sigma(1), \dots, \sigma(k)$ , and  $\sigma(k+1) = i$ .
- (b) The  $\{\sigma(1), \sigma(k+1), \dots, \sigma(n)\}$ -tree derived from  $x$  is the same as the  $\{\sigma(1), \sigma(k+1), \dots, \sigma(n)\}$ -tree derived from  $v$ .
- (c) The  $\{\sigma(1), \dots, \sigma(k+1), \sigma(n)\}$ -tree derived from  $y$  is the same as the  $\{\sigma(1), \dots, \sigma(k+1), \sigma(n)\}$ -tree derived from  $w$ .

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  $\sigma(1), \dots, \sigma(k)$ . Event (b) must occur because none of the leaves  $\sigma(1), \sigma(k+1), \dots, \sigma(n)$  is moved in the first  $k-1$  steps in the construction of  $\gamma_{xy}$ , so the  $\{\sigma(1), \sigma(k+1), \dots, \sigma(n)\}$ -tree derived from  $x$  is the  $\{\sigma(1), \sigma(k+1), \dots, \sigma(n)\}$ -tree derived from  $x_k$ . Also, event (c) must occur because the leaves  $\sigma(1), \dots, \sigma(k+1), \sigma(n)$  are not moved again after the  $k$ th step in the construction, so the  $\{\sigma(1), \dots, \sigma(k+1), \sigma(n)\}$ -tree

derived from  $y$  is the same as the  $\{\sigma(1), \dots, \sigma(k+1), \sigma(n)\}$ -tree derived from  $x_{k+1}$ .

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

$$\begin{aligned} & \sum_{x \in V} \sum_{y \in V} \pi_x \pi_y P(e \in \gamma_{xy} \text{ at step } k) \\ & \leq \frac{1}{\binom{n}{k} (n-k) c_{k+2} c_{n-k+1}} \\ & = \frac{1}{\binom{n}{k} (n-k) (2k-1) (2k-3) (2(n-k)-3) c_k c_{n-k}} \end{aligned} \quad (7)$$

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 (5) and (7), we get

$$\tau_n \leq 16A_1 n^{7/2} \max_{e \in E} \max_{k \in K(e)} \frac{c_n}{\binom{n}{k} (n-k) (2k-1) (2k-3) (2(n-k)-3) c_k c_{n-k}}. \quad (8)$$

By Stirling's formula, we have

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

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

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

Also, we have

$$\begin{aligned} \binom{n}{k} &= \frac{n!}{k!(n-k)!} \approx \frac{n^{n+1/2} e^{-n}}{(k^{k+1/2} e^{-k})((n-k)^{n-k+1/2} e^{-(n-k)})} \\ &= \frac{n^{n+1/2}}{k^{k+1/2} (n-k)^{n-k+1/2}}, \end{aligned}$$

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  $k \in \{1, \dots, n - 1\}$ , we have

$$\begin{aligned} & \frac{c_n}{\binom{n}{k}(n-k)(2k-1)(2k-3)(2(n-k)-3)c_k c_{n-k}} \\ & \leq \frac{A_3}{k^2(n-k)^2} \frac{k^{k+1/2}(n-k)^{n-k+1/2}}{n^{n+1/2}} \frac{2^n n^{n-2} e^{-n}}{(2^k k^{k-2} e^{-k})(2^{n-k}(n-k)^{n-k-2} e^{-(n-k)})} \\ & = \frac{A_3}{k^2(n-k)^2} \left( \frac{k(n-k)}{n} \right)^{5/2} = A_3 n^{-5/2} (k(n-k))^{1/2} \leq A_3 n^{-3/2} \end{aligned} \tag{9}$$

for all  $e$  and all  $k \in K(e)$ . Thus, from (8) and (9), we have  $\tau_n \leq 16A_1 A_3 n^2$ . Hence,  $\tau_n = O(n^2)$ , as claimed.

#### 4. CONCLUDING REMARKS

Although it would be easy to sample directly from the set of  $n$ -leaf cladograms (see the proof of Lemma 4), Markov chains on other combinatorial sets have been studied for the purpose of obtaining fast algorithms for generating approximately uniform samples from these sets. For example, Randall and Tetali [14] studied Markov chains on lozenge tilings, domino tilings, triangulations of convex polygons, and binary trees. A chain on  $m \times n$  contingency tables was studied in [7], and chains on independent sets of a graph were studied in [6, 12]. However, the upper bounds obtained for the mixing times of these chains are not known to be sharp. The chain in this article provides an example of a chain on a combinatorial set for which the relaxation time is known up to a constant factor.

Regarding the techniques used in this article, the advantage that we get from using Corollary 3, rather than one of the standard Poincaré inequalities, is that we only need to bound the bottlenecking for the paths that end in the set  $B$ . Consequently, we do not need to consider the paths ending in trees which are unusually long and thin. Although the author is not currently aware of other useful applications of Corollary 3, it seems likely that there would be other Markov chains for which one could obtain a better bound by avoiding having to consider paths which end in a “bad” part of the state space.

The Markov chain in this article appears to be the easiest Markov chain on cladograms to analyze. A natural next step might be to consider other chains on  $n$ -leaf cladograms. For example, Aldous [3] describes a chain in which each transition consists of picking a random edge (not necessarily adjacent to a leaf), removing the entire branch of the chain below that edge, and then reattaching the branch to one of the edges adjacent to where it was before. Aldous poses the problem of proving that this chain has a relaxation time of  $O(n^{3/2})$ . He also mentions that this chain is believed to be related to a chain on triangulations of the regular  $n$ -gon.

#### ACKNOWLEDGMENTS

The author thanks 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 article. He also thanks the two referees for their comments which led to improvements in the exposition of the article.

## REFERENCES

- [1] D. Aldous, The continuum random tree III. *Ann. Probab.* 21 (1993), 248–289.
- [2] D. Aldous, Mixing time for a Markov chain on cladograms. *Combin. Probab. Comput.* 9:191–204, 2000.
- [3] D. Aldous, My favorite 6 open problems. Unpublished manuscript. Available via <http://www.stat.berkeley.edu/~aldous>, 2000.
- [4] D. Aldous and J. Fill, Reversible Markov chains and random walks on graphs. Book in preparation. Available via <http://www.stat.berkeley.edu/~aldous>.
- [5] P. Diaconis and D. Stroock, Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* 1 (1991), 36–61.
- [6] M. Dyer and C. Greenhill, On Markov chains for independent sets. *J. of Algorithms*, 35 (2000), 17–49.
- [7] M. Dyer, R. Kannan, and J. Mount, Sampling contingency tables. *Random Structures Algorithms* 10 (1997), 487–506.
- [8] J. Fulman and E. Wilmer, Comparing eigenvalue bounds for Markov chains: When does Poincaré beat Cheeger? *Ann. Appl. Probab.* 9 (1999), 1–13.
- [9] M. Jerrum and A. Sinclair, Approximating the permanent. *Siam J. Comput.* 18 (1989), 1149–1178.
- [10] N. Kahale, A semidefinite bound for the mixing rates of Markov chains. *Random Structures Algorithms* 11 (1997), 299–313.
- [11] L. Lovász and P. Winkler, Mixing times, *Microsurveys in Discrete Probability*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 41, D. Aldous and J. Propp (Editors), American Mathematical Society, 1988, pp. 85–134.
- [12] M. Luby and E. Vigoda, Fast convergence of the Glauber dynamics for sampling independent sets. *Random Structures Algorithms* 15 (1999), 229–241.
- [13] J. Pitman and M. Yor, The law of the maximum of a Bessel bridge. *Electron. J. Probab.*, 4 (1999), 1–35.
- [14] D. Randall and P. Tetali, Analyzing Glauber dynamics by comparison of Markov chains. *J. Math. Phys.* 41 (2000), 1598–1615.
- [15] A. Sinclair, Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.* 1 (1992), 351–370.