

# Linear cover time for trees is exponentially unlikely

Amir Yehudayoff\*

*Received: November 8, 2011; revised: July 25, 2012; published: July 28, 2012.*

**Abstract:** The cover time of a graph is the time it takes for a random walk on it to visit all vertices. This note shows that for any tree on  $n$  vertices, the probability that the cover time of the tree is linear in  $n$  is exponentially small in  $n$ . This is true for any starting vertex.

## 1 Introduction

The cover time of a graph is the time it takes for a random walk on it to visit all its vertices. One well-known example of a cover time question is the coupon collector problem: How fast can a coupon collector, who gets random coupons, collect all possible coupons? This problem corresponds to the cover time of the complete graph or the star graph (the tree with all vertices but one of degree one).

It is a basic property of graphs that has been studied for more than thirty years, and has various uses and applications [1, 5, 9]. For general graphs, the only statement known was proved by Feige [5]: A  $(1 - o(1))n \ln n$  lower bound on the expected cover time of graphs of size  $n$ . Benjamini, Gurel-Gurevich and Morris have recently proved that for graphs of constant degree, the probability that the cover time is linear is exponentially small [3]. A similar probability estimate holds for expander graphs (follows, e.g., by Gilman's Chernoff-type estimate for expanders [6]).

Here we consider the cover time of trees (simple connected graphs with no cycles). Previously, Brightwell and Winkler [4] proved an order  $n \ln n$  lower bound on the expected cover time of trees of size  $n$ . Two recent works [7, 8] study other aspects of cover time of trees. Our main result is that not only that the expected cover time of a tree is super-linear, but the probability that it is linear is exponentially small. In other words, the probability of collecting all coupons on a tree in linear time is extremely small.

---

\*Horev fellow – supported by the Taub Foundation. Research supported by ISF and BSF.

**Key words and phrases:** cover time, trees, coupon collector

For a vertex  $v$  in a tree  $T$  of size  $n$ , denote by  $\Pr_v$  the probability with respect to a simple random walk  $\{X_t\}_{t=0}^\infty$  on  $T$  started at  $v$ . Define the *cover time* as the random variable

$$\tau = \min\{t \geq 0 : |\{X_0, X_1, \dots, X_t\}| = n\}.$$

**Theorem 1.1.** There exists a constant  $C > 0$  so that the following holds. For all trees  $T$  of size  $n$ , for all  $v$  in  $T$ , and for all  $L > 0$ ,

$$\Pr_v[\tau \leq Ln] \leq C2^{-n2^{-cL}}.$$

The estimate is optimal in the sense that a similar lower bound on the probability holds for the star graph.

Nonaka et al. [8] has shown that the tree that minimizes the expected cover time is the star. It would perhaps be natural to guess that the tree that maximizes the probability of a linear cover time is the tree as well. This, however, is not the case. Consider, for example, the two graphs:  $s_n$  the star on  $n$  vertices and  $\ell_n$  the line graph on  $n$  vertices. The probability of covering  $s_n$  in time at most  $2n$  is roughly  $n!/n^n \sim e^{-n}$ . The probability to cover  $\ell_n$  in time at most  $2n$  is much larger, roughly  $2^{-2n} \binom{2n}{1.5n} \sim 2^{-0.4n}$ .

The proof is based on two simple observations:

- (i) For all integers  $s$ , every tree  $T$  of size  $n$  can be partitioned to order  $n/s$  edge-disjoint sub-trees  $T_1, \dots, T_N$ , each of size order  $s$ . For a formal statement, see Lemma 2.1 below.
- (ii) Given a random walk on  $T$ , we can consider the “marginal” of the random walk on one of the sub-trees  $T_i$ , that is, the part of the walk that is performed only inside  $T_i$ . The observation is that these  $N$  marginals are “independent.” For example, if  $Y = (Y_1, \dots, Y_i)$  is the first  $i$  positions of the walk in  $T_1$ , and  $Z = (Z_1, \dots, Z_j)$  is the first  $j$  positions of the walk in  $T_2$ , then  $Y$  and  $Z$  are independent. This is true since if we leave  $T_i$  at a vertex  $u$  we must return to  $T_i$  at the same vertex  $u$ . For more details, see Claim 2.2 below.

The two observations are used twice: Once to get a non-trivial estimate and once to get the exponential estimate. First, use (ii) to (roughly) bound the variance of the cover time, and deduce that for all  $i \in [N]$ , the probability that the cover time of  $T_i$  is less than  $Ls$  is at most  $1/2$ . For this, use the cover time expectation estimate of Brightwell and Winkler [4], and choose  $s$  exponential in  $L$ . Second, use the non-trivial estimate and (ii) again to deduce that the probability that the number of  $T_i$ 's that are covered in less than  $Ls$  steps is less than  $N/3$  is exponentially small. Under the complement of this low probability event, the cover time of  $T$  is at least order  $NLs = nL$ .

## 2 Bounding the cover time from below

### 2.1 Partitioning trees

The following lemma describes how to equipartition a tree to sub-trees.

**Lemma 2.1.** *Let  $T$  be a tree of size  $n \geq 2Ns$ , for  $N, s$  positive integers. Then, there exists  $T_1, \dots, T_N$  edge-disjoint sub-trees of  $T$  so that  $s \leq |T_i| \leq 2s$  for every  $i \in [N]$ .*

*Proof.* Prove the lemma by induction on  $N$ . For  $N = 1$ , the lemma trivially holds. Assume  $N > 1$ .

Find a sub-tree  $T_1$  of  $T$  of size  $s \leq |T_1| \leq 2s$ . To see that this is always possible, choose an arbitrary root  $v_0$  for  $T$  and direct all edges toward  $v_0$ . For every node  $v$  in  $T$ , denote by  $T_v$  the sub-tree of  $T$  rooted at  $v$  (with all edges directed toward  $v$ ). Observe  $|T_{v_0}| > s$ . Let  $u$  be a vertex so that  $|T_u| > s$  but  $|T_v| \leq s$  for every  $v$  in  $T_u$ . Let  $w_1, \dots, w_m$  be the children of  $u$ , that is, the edge  $(u_i, u)$  exists in  $T$  for all  $i \in [m]$ . There, thus, exists  $W \subseteq \{w_1, \dots, w_m\}$  so that the size of the sub-tree of  $T$  induced on  $\{u\} \cup \bigcup_{w \in W} V(T_w)$  is at least  $s$  and at most  $2s$ , where  $V(T_w)$  are the vertices of  $T_w$ . Set  $T_1$  as this induced sub-tree.

Set  $T_2$  as the connected component of  $u$  in the tree  $T$  after removing all edges in  $T_1$ . The size of  $T_2$  is at least  $n - |T_1| \geq 2(N - 1)s$ . Apply the induction hypothesis on  $T_2$  to get a collection of  $N - 1$  sub-trees of  $T_2$  with the appropriate properties. Adding  $T_1$  to the collection of sub-trees completes the proof.  $\square$

## 2.2 Using the partition to get independence

Fix a tree  $T$  of size  $n$ , fix a start vertex  $v$  and fix an integer  $L \geq 1$ . Let  $C > 0$  be a universal constant to be determined. Apply Lemma 2.1 to  $T$  with  $s = \lceil 2^{CL} \rceil$  to obtain  $N = \lfloor n/(2s) \rfloor$  edge-disjoint sub-trees  $T_1, \dots, T_N$  of  $T$  of size order  $s$ . Denote by  $v_i$  the unique vertex in  $T_i$  that is first on the path from  $v$  to  $T_i$ . It could be the  $v_i$  is  $v$ .

Let  $\sigma_i$  be the number of steps the random walk makes along edges in  $T_i$  before covering  $T_i$  (which typically happens prior to covering  $T$ ). Denote

$$\sigma = \sigma_1 + \dots + \sigma_N. \quad (2.1)$$

A.s.,  $\sigma$  bounds the cover time  $\tau$  of  $T$  from below.

The following claim is the main observation behind the proof.

**Claim 2.2.** *The random variables  $\sigma_1, \dots, \sigma_N$  are independent. Each  $\sigma_i$  is distributed as the cover time of the tree  $T_i$  with start vertex  $v_i$ .*

*Proof.* Fix  $i \in [N]$ , and focus on  $T_i$  and  $\sigma_i$ . We argue that, conditioned on the values of  $\{\sigma_j : j \neq i\}$ , the random variable  $\sigma_i$  is distributed as the cover time of the tree  $T_i$  with start vertex  $v_i$ .

Start by slightly changing the way the random walk on  $T$  is sampled. For every  $u$  in  $T_i$  so that there is an edge from  $u$  to outside of  $T_i$ , define an infinite sequence of i.i.d. zero-one random variables  $\{\xi_t(u)\}$  with  $\Pr[\xi_t(u) = 1] = d_i(u)/d(u)$ , where  $d_i(u)$  is the degree of  $u$  inside  $T_i$  and  $d(u)$  is the degree of  $u$  in  $T$ . When  $X_t = u$  for such a  $u$ , the distribution of  $X_{t+1}$  is as follows. If  $\xi_t(u) = 1$ , then  $X_{t+1}$  is a uniform neighbor of  $u$  in  $T_i$ . If  $\xi_t(u) = 0$ , then  $X_{t+1}$  is a uniform neighbor of  $u$  outside  $T_i$ . The random walk on other vertices is defined as usual.

The random variable  $\sigma_i$  is distributed as claimed, even when conditioned on more than just the values of  $\{\sigma_j : j \neq i\}$ . Condition on the values of  $\xi_t(u)$  for all  $u, t$  and on all choices the random walk makes when at vertices not in  $T_i$ . This fixes the value of  $\sigma_j$  for all  $j \neq i$ . The first step the random walk makes in  $T_i$  is always at  $v_i$ , as it starts at  $v$ . Moreover, since  $T$  has no cycles and  $T_i$  is connected, when the random walk leaves  $T_i$  at a vertex  $u$ , the only place where it can come back to  $T_i$  is at  $u$ . For every such conditioning (a.s.), therefore, the distribution of  $\sigma_i$  is that of the cover time according to a random walk on  $T_i$  (there is some “deterministic time shift of the steps in  $T_i$ ” that does not affect  $\sigma_i$ ).  $\square$

### 2.3 A non-trivial estimate using second moment

First, use independence to prove the following non-trivial probability estimate.

**Proposition 2.1.** Let  $T$  be a tree of size  $n$  and let  $v$  be a vertex in it. Let  $\tau$  be the cover time of  $T$  and let  $\sigma$  be as defined in (2.1). If  $n \geq 2^{CL}$ , then

$$\Pr_v[\tau \leq Ln] \leq \Pr_v[\sigma \leq Ln] \leq \frac{\text{Var}_v(\sigma)}{L^2 n^2} \leq 1/2$$

with  $C > 0$  a universal constant.

The proposition follows from the following two lemmas, using Chebyshev's inequality.

**Lemma 2.3.**  $\mathbb{E}_v[\sigma] \geq Ns \log(s)/C$  with  $C > 0$  a universal constant.

**Lemma 2.4.**  $\text{Var}_v(\sigma) \leq Cns$  with  $C > 0$  a universal constant.

To prove the expectation lower bound, use the following theorem of Brightwell and Winkler [4]. They showed that the tree minimizing the expectation of the cover time is the star. Feige [5] proved a similar asymptotic result that holds for general graphs.

**Theorem 2.2.** Let  $T'$  be a tree of size  $n'$  and let  $v'$  be a vertex in it. Let  $\tau'$  be the cover time of  $T'$ . Then<sup>1</sup>,

$$\mathbb{E}_{v'}[\tau'] \geq n' \log(n')/C'$$

with  $C' > 0$  a universal constant.

*Proof of Lemma 2.3.* Since  $|T_i| \geq s$  for all  $i \in [N]$ , Claim 2.2 and Theorem 2.2 tell us that  $\mathbb{E}_v[\sigma] \geq Ns \log(s)/C$ .  $\square$

To prove the variance upper bound, use the following well-known claim [2].

**Claim 2.5.** Let  $T'$  be a tree of size  $n'$  and let  $v'$  be a vertex in it. Let  $\tau'$  be the cover time of  $T'$ . Then,

$$\mathbb{E}_{v'}[\tau'] \leq C'n'^2$$

with  $C' > 0$  a universal constant.

Claim 2.5 has the following corollary.

**Corollary 2.3.** Let  $T'$  be a tree of size  $n'$  and let  $v'$  be a vertex in it. Let  $\tau'$  be the cover time of  $T'$ . Then,  $\Pr_{v'}[\tau' \geq 2C'n'^2] \leq 1/2$ . Thus, by Markov's property,  $\Pr_{v'}[\tau' \geq 2C'kn'^2] \leq (1/2)^k$  for all integers  $k$ . Finally,

$$\text{Var}_v(\tau') \leq \mathbb{E}_{v'}[\tau'^2] \leq C''n'^2$$

with  $C'' > 0$  a universal constant.

---

<sup>1</sup>Logarithms are base two.

*Proof of Lemma 2.4.* By Claim 2.2,

$$\text{Var}_v(\sigma) = \text{Var}_v(\sigma_1) + \dots + \text{Var}_v(\sigma_N).$$

Using Corollary 2.3 as well, since  $|T_i| \leq 2s$  for all  $i \in [N]$ ,

$$\text{Var}_v(\sigma) \leq NC''(2s)^2 \leq Cns.$$

□

## 2.4 An exponential estimate

The non-trivial estimate together with the independence imply the tight estimate.

*Proof of Theorem 1.1.* Define the random variables

$$\xi_i = \begin{cases} 1 & \text{if } \sigma_i \geq 8Ls, \\ 0 & \text{otherwise.} \end{cases}$$

Denote

$$\xi = \xi_1 + \dots + \xi_N.$$

Since  $s \geq 2^{CL}$ , by Proposition 2.1 applied on the tree  $T_i$ , and by Claim 2.2, a.s.

$$\Pr_v[\xi_i = 1 | \xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_N] \geq 1/2.$$

So,

$$\Pr_v[\xi \leq N/3] \leq 2^{-N/C}.$$

As  $T_1, \dots, T_N$  are edge-disjoint, a.s.

$$\tau \geq 8Ls\xi.$$

Hence,

$$\Pr[\tau \leq Ln] \leq \Pr[\xi \leq n/(8s) \leq N/3].$$

□

## References

- [1] D. J. Aldous. Threshold limits for cover times. *Journal of Theoretical Probability* 4, pages 197–211, 1991. 1
- [2] D. Aldous and J. Fill. Reversible Markov chains and random walks on graphs. Manuscript available at <http://www.stat.berkeley.edu/~aldous/RWG/book.html>, 1999. 4
- [3] I. Benjamini, O. Gurel-Gurevich and B. Morris. Linear cover time is exponentially unlikely. arXiv:1011.3118v1, 2010. 1

- [4] G. Brightwell and P. Winkler. Extremal cover time for random walks on trees. *Journal of Graph Theory* 14 (5), pages 547–554, 1990. [1](#), [2](#), [4](#)
- [5] U. Feige. A tight lower bound on the cover time for random walks on graphs. *Random Structures and Algorithms* 6 (4), pages 433–438, 1995. [1](#), [4](#)
- [6] D. Gillman. A Chernoff bound for random walks on expander graphs. *SIAM Journal on Computing* 27 (4), pages 1203–1220, 1998. [1](#)
- [7] Y. Nonaka, H. Ono, K. Sadakane and M. Yamashita. How to design linear cover time random walk on a finite graph. *Stochastic Algorithms: Foundations and Applications, Lecture Notes in Computer Science* 5792, pages 104–116, 2009. [1](#)
- [8] Y. Nonaka, H. Ono, S. Kijima and M. Yamashita. How slow, or fast, are standard random walks? analyses of hitting and cover times on tree. *Australian Computer Science Communications* 33 (8), pages 63–68, 2011. [1](#), [2](#)
- [9] D. Zuckerman. A technique for lower bounding the cover time. *SIAM J. Disc. Math*, pages 81–87, 1992. [1](#)

## AUTHOR

Amir Yehudayoff  
Technion–IIT, Haifa, Israel  
amir.yehudayoff@gmail.com  
<http://www.technion.ac.il/~yehuday/>