Multiple Cover Time

Peter Winkler
Bell Laboratories, Murray Hill, NJ 07974

David Zuckerman
Department of Computer Sciences, UT Austin, Austin, TX 78712

Received 14 June 1996; Accepted 19 July 1996

ABSTRACT: Motivated by applications in Markov estimation and distributed computing, we define the blanket time of an undirected graph $G$ to be the expected time for a random walk to hit every vertex of $G$ within a constant factor of the number of times predicted by the stationary distribution. Thus the blanket time is, essentially, the number of steps required of a random walk in order that the observed distribution reflect the stationary distribution. We provide substantial evidence for the following conjecture: that the blanket time of a graph never exceeds the cover time by more than a constant factor. In other words, at the cost of a multiplicative constant one can hit every vertex often instead of merely once. We prove the conjecture in the case where the cover time and maximum hitting time differ by a logarithmic factor. This case includes almost all graphs, as well as most "natural" graphs: the hypercube, $k$-dimensional lattices for $k \geq 2$, balanced $k$-ary trees, and expanders. We further prove the conjecture for perhaps the most natural graphs not falling in the above case: paths and cycles. Finally, we prove the conjecture in the case of independent stochastic processes. © 1996 John Wiley & Sons, Inc. Random Struct. Alg., 9, 403–411 (1996)

Key Words: random walk; cover time; blanket time; graph; path; cycle; coupon collecting
1. INTRODUCTION

A random walk on a connected, undirected graph $G$ with $n$ vertices is a Markov chain whose states are the vertices of $G$. The walk begins with a token at some vertex $v$, and at each tick of the clock, the token moves with equal probability to any vertex adjacent to its current position. If instead the transition probabilities are biased according to edge weights, one obtains a general reversible Markov chain (to which most of the following remarks also apply).

Random walks on graphs have proved to be a useful tool in several aspects of the theory of computing, as well as an elegant subject for mathematical analysis. The most notable, perhaps, of recent applications has been to polynomial-time randomized approximation algorithms (e.g., [16, 10, 12]), but there have also been uses in space complexity [4], online algorithms [9], and distributed computing [15].

There are at least three graph parameters of the form “expected number of steps of a random walk before $X$ happens” that have been extensively studied. The mixing time is the least $t$ such that the distribution of the walk at step $t$ is approximately (within variation distance $1/2$ of) the stationary distribution; the maximum hitting time, denoted by $H$, is the expected time to reach $u$ from $v$, maximized over $u$ and $v$; the maximum cover time, denoted by $C$, is the expected time to hit all vertices, maximized over starting vertex.

To these we add a fourth parameter, larger than the others, which is described roughly as the expectation of the least $t$ such that by step $t$ the walk has hit every vertex about the expected number of times predicted by the stationary distribution. We call this parameter the “blanket time,” denoted by $B$. Thus, for a regular graph, $B$ is the least positive $t$ before all vertices are covered $\Omega(t/n)$ times. The blanket time is quite different from the mixing time in that it requires the observed data from a single walk to reflect the stationary distribution.

In studying blanket time we are motivated by several problems arising in computer science. First, the vertices (states) of an unknown random walk (resp. Markov chain) may be observed with the object of estimating the stationary distribution. The blanket time is an indication of how many observations must be made. An online algorithm, for example, may attempt to organize a data structure in accordance with a presumed stationary distribution for a Markov input model; the blanket time then indicates how long it takes for the process to settle down. This is also relevant to the well-studied and more difficult program of inferring a Markov chain from the labels of the edges traversed [5, 19].

Blanket time arises naturally also in randomized distributed algorithms of theoretical interest, such as token management systems proposed by several authors [6, 15]. Suppose that a token (representing, e.g., privileged access to a critical section) is passed randomly from processor to processor in a network: How long will it be before every processor has owned the token about as long as it is supposed to? Even if the network is not synchronous, the imposition of a fairness condition requiring processors to be active approximately equally often allows us to answer the above question using blanket time.

In these applications, one obviously needs to wait at least until all states (processors) have been hit; in other words, $B \geq C$. However, we have come to believe that the following striking converse is true: $B = O(C)$. If true this would of course strengthen all of the many results bounding cover time for various classes of
graphs: Not only is each vertex visited once, but each is visited the “right” number of times.

We support this conjecture with three results. To motivate the first, one very useful bound on the cover times is $C = O(H \log n)$ [1, 18]. We strengthen this by showing that in fact $B = O(H \log n)$. This establishes our conjecture in the very common case where $C = \Theta(H \log n)$. This is the case for almost all graphs, as well as most “natural” graphs: The hypercube [1], $k$-dimensional lattices for $k \geq 2$ [1, 24], balanced $k$-ary trees [23, 11, 24], and expanders [7, 20, 6].

The “classic” examples of graphs for which $C = \Theta(H)$ are paths and cycles. In our second result, we show that for these graphs our conjecture also holds. Aldous [3] has subsequently shown that this second theorem also follows from deep but well-known facts about Brownian motion [22]. Our proof, on the other hand, is elementary. Neither proof, however, appears easy to generalize.

Finally, we prove the conjecture in the case of independent stochastic processes (current state not dependent on previous state). It turns out that we are able to prove a theorem (Theorem 1 below) that simultaneously handles this case and the case where $C = \Theta(H \log n)$.

2. DEFINITIONS AND NOTATION

The random walk is performed on a connected, undirected graph $G(V, E)$. We define the following, generally using boldface to denote random variables:

- $n = |V|$, $\pi_v = \deg(v)/2|E|$ is the stationary probability at vertex $v$,
- $E_v(\cdot)$ is the expected value of $\cdot$ in a random walk starting at vertex $v$,
- $H_v$ is the time to first visit vertex $v$,
- $H_v = \max_{u \in V} \{E_u H_v\}$,
- $H = \max_{u \in V} \{E_u H_v\}$,
- $C$ is the cover time, i.e., $\min \{t : (\forall v)H_v \leq t\}$,
- $C$ is the cover time, i.e., $\min \{t : (\forall v)H_v \leq t\}$,
- $N_v(t)$ is the number of time vertex $v$ is visited by time $t$.

We can now define blanket time, and our conjecture. Blanket time will come with another parameter $\delta$, which represents the ratio of the number of times $v$ must be hit to the number that it should be hit.

**Definition.** For $\delta \in [0, 1)$, define the random variable $B_\delta = \min \{t : (\forall v)N_v(t) > \delta \pi_v t\}$. The blanket time $B_\delta = \max_{v \in V} E_v B_\delta$.

Note that $B_0 = C$ denotes the ordinary cover time (because of the strict inequality). Our conjecture may now be stated:

**Conjecture.** There exists $\delta, a > 0$ such that for all graphs, $B_\delta \leq aC$.

This is equivalent to saying that the expected time until each vertex $v$ is visited $\pi_v C$ times is $O(C)$. 
There is also a strong form of our conjecture, which we also prove in our special cases except for the case of paths and cycles:

**Conjecture (Strong Form).** For all \( \epsilon > 0 \), there exists \( a_\epsilon > 0 \) such that for all graphs, \( B_{1-\epsilon} \leq a_\epsilon C \).

### 3. CONJECTURE IN THE CASES \( C = \Theta(H \log n) \) AND INDEPENDENT STOCHASTIC PROCESSES

The first special case on which one might wish to test the blanket time conjecture is the Markov chain in which all transition probabilities are equal; that is, a sequence of independent, discrete, uniform random variables. The cover time in this case is the solution to the famous "coupon collector's problem." If there are \( n \) coupons to be collected, then it is easy to see that the expected time to collect at least one of each is

\[
\sum_{i=1}^{n} \frac{n}{n-i+1} \sim n \log n.
\]

How much longer does it take to get a representative collection? In [17] the answer can be found as a consequence of Lemma 11, (Chap. 2, Section 6, p. 113): If \( cn \ln n \) coupons are collected, where \( c \) is a constant greater than 1, then the least-collected coupon will be represented by about \( \delta c \ln n \) coupons, where \( \delta \) is the positive root of

\[
\delta + c(\ln \delta - \delta + 1) = 0.
\]

It follows that for coupon collecting, the blanket time \( B_\delta \) is asymptotically \( cn \ln n \), where \( c = \frac{\delta}{\delta - \ln \delta - 1} \), provided \( \delta \) is large enough so that \( c > 1 \). For example, \( B_{1/2} \sim 2.59n \ln n \), so that at the cost of waiting 2.59 times as long as necessary for a single complete collection, we can have half the number of complete collections possible in the given time.

The case of nonuniform i.i.d. random variables, i.e., nonequiprobable coupons, is considered in [17] as well as other articles on coupon collecting (see, e.g., [13, 21, 14]). Although the cover time problem is solved, the literature does not appear to supply the means for computing blanket time in the non-uniform case.

However, using a result of Aldous [2], we are able to derive the result for independent stochastic processes from our theorem below, which covers Markov chains in which \( C = \Theta(H \log n) \).

Define \( S \) as the real number satisfying

\[
\sum_{v \in V} e^{-S/H_v} = e^{-1}.
\]

We first show

**Lemma 1.** \( C < 2eS \).

**Proof.** We extend the proof in [1, 23] that \( C = O(H \log n) \). For any vertex \( u \),
Pr_u[H_v \geq eH_v] \leq 1/e, so for integral k, Pr_u[H_v \geq ekH_v] \leq e^{-k}. Thus

\[ \Pr_u[C \geq ceS] \leq \sum_{v \in V} e^{-e^{S/H_v}} \leq e^{-\epsilon S/H_v} \leq e^{-\epsilon^{(x-1)}}. \]

Thus \( E_u[C/(eS)] \leq 1 + \int_{x-1}^\infty e^{-\epsilon^{(x-1)}}

We now prove the much stronger theorem

**Theorem 1.** \( B_{1-\epsilon} = O(S/e^2). \)

Since \( S \leq H \log n \), this implies

**Corollary 1.** \( B_{1-\epsilon} = O(H \log n/e^2). \)

To move the case of independent stochastic processes, i.e., where there is no dependence on previous state, we make use of a known relationship between the cover time and the parameter \( S \) defined above, namely, that \( C = \Theta(S) \) (viz. Aldous [2]). This gives

**Corollary 2.** For independent stochastic processes, \( B_{1-\epsilon} = O(C/e^2). \)

Note also that if there is a constant number of hard-to-hit vertices, then \( S \) and \( C \) are both \( \Theta(H) \), so the conjecture is proved in this case.

**Proof of Theorem 1.** The outline of the proof is to show that any fixed vertex is visited the appropriate number of times with high probability. To prove this, we use the fact that, although it may take a long time to reach a vertex, once we do, it takes the “appropriate” amount of time to visit it again.

To simplify notation, we sometimes assume when it is insignificant that numbers are integers; moreover, we obtain a bound for \( B_{1-\epsilon} \).

First walk for \( C \) steps, i.e., until all vertices are hit. Since \( C = O(S) \), it suffices to show that for some constant \( a \geq 1/e \), if we walk for an additional \( a(1 + \epsilon)S \) steps, with high probability all vertices are covered at least \( \pi_v A S \) times.

Let \( H_v^j \) be the time at which vertex \( v \) has been visited \( j \) times, and let \( k_v = H_v/\pi_v \). Then \( E_v(H_v^k_v) = k_v/\pi_v = H_v \). Moreover, for any other vertex \( u \),

\[ E_u H_u^{k_v} \leq H_v + (k_v - 1)/\pi_v < 2H_v. \]

So

\[ \Pr_u[H_v^{k_v} \geq 2eH_v] < 1/e. \]

Since \( u \) was arbitrary,

\[ \Pr_u[H_v^{k_v} \geq 2qeH_v] < e^{-q}. \]
In order to apply the Chernoff bound, we define $X = \frac{H}{4e^H}$. Then $EX = 1/4e$, and

$$E[e^X] = \int_{y=0}^{\infty} \Pr[X \geq y] dy = \int_{y=0}^{\infty} \Pr[X \geq \ln y] dy \leq 1 + \int_{y=1}^{\infty} e^{-[2\ln y]} dy \leq 1 + \int_{y=1}^{\infty} \frac{e}{y^2} dy = 1 + e < 4.$$  

We now wish to add independent copies of $Y = X - (1 + \epsilon)EX$, and show that the sum is unlikely to be positive. To do this, it suffices to find a $t > 0$ that makes $E[e^{tY}] < 1$. This follows from Chernoff [8]; however, we proceed with the calculation in order to find the dependence on $\epsilon$. Using the Taylor series expansion, it is not hard to show that for $0 < t < 1$,

$$E[e^{tY}] < 1 + tEY + (t^2/2)E[e^Y - 1 - Y] < 1 - t/4e + 2t^2.$$  

This last expression attains a minimum value $1 - \epsilon^2/c$, where $c = 128e^2 < 1000$. Therefore, by Chernoff, if $Y_b$ is the sum of $b$ independent copies of $Y$, then

$$\Pr[Y_b \geq 0] \leq (1 - \epsilon^2/c)^b < e^{-\epsilon^2b/c}. \quad (1)$$  

To translate this statement back to multiple hitting times for $v$, note that the sum of $b$ copies of $H_v^k$ is equal to $H_v^{bk}$. Thus (1) becomes

$$\Pr[H_v^{bk} \geq (1 + \epsilon)bk_v/\pi_v] < e^{-\epsilon^2b/c}.$$  

Thus, taking $b = dcS/H_v, \epsilon^2$ for some constant $d$, we get $bk_v = dcS\pi_v/e^2$, and

$$\Pr[(\exists v)H_v^{bk} \geq dcS/e^2] \leq \sum_{v \in V} e^{-dS/H_v} \leq e^{-(d-1)S/H} \sum_{v \in V} e^{-S/H_v} \leq e^{-(d-1)},$$  

since $S \geq H$. The theorem follows.

4. CONJECTURE FOR THE CYCLE AND PATH

It is clear that if the conjecture holds for the cycle, then it holds also for the path (note that we will not prove the conjecture in the strong form).

Let $G_n$ be the cycle on vertices $0, 1, \ldots, n - 1$ with the vertices regarded modulo $n$. Let $C_n^{[A]}$ denote the least $t$ such that a particular random walk on $G_n$ beginning at 0 has hit each vertex $\lambda$ times by time $t$, and lightface $C_n^{[A]}$ denotes its expectation. Thus our conjecture for the cycle is equivalent to:

**Theorem 2.** $C_n^{[n]} = O(n^2)$.

**Proof.** We show by induction on $k$ that

$$C_n^{[A]} \leq F_n \lambda n,$$
where $\lambda \geq n = 2^k$, $F_2 = 1$, and $F_j = (1 + 9j^{-1/4})F_{j/2}$ for $j > 2$. To see that this would suffice to prove the theorem, note first that

$$\ln(F_{2^i}) = \sum_{i=1}^{k-2} \ln(1 + 9 \cdot 2^{-i/4}) < 9 \sum_{i=1}^{k-2} (2^{-1/4})^i < \frac{9}{1 - 2^{-1/4}} < 48,$$

so $F_n$ is bounded by $e^{48}$. We then have for any $n$, not necessarily a power of 2, that $C_n^{[\lambda]} \leq C_n^{[n]} \leq e^{48}(n')^2 < 4e^{48}n^2$, where $n'$ is the least power of 2 greater than or equal to $n$.

We base the induction at $n = 2$, where $C_n^{[\lambda]} = 2\lambda - 1 < F_2 \lambda n$. Let us now fix $n = 2^k$ and $\lambda \geq n$, so that we may assume

$$C_n^{[A/2 + \lambda^{3/4}]} \leq F_n^{[A/2 + \lambda^{3/4}]} \cdot \frac{(\lambda/2 + \lambda^{3/4})}{(n/2)}.$$

Our approach will be to couple a random walk $W$ on $G_n$ with a random walk $W'$ on $G_{n/2}$ as follows: Both walks begin at $0$; whenever $W$ steps $2i$ to $2i + 1$ to $2i + 2$, $W'$ steps from $i$ to $i + 1$; whenever $W$ steps $2i$ to $2i - 1$ to $2i - 2$, $W'$ steps from $i$ to $i - 1$. In other words, we treat the even vertices of $G_n$ as a copy of $G_{n/2}$.

Let $T'_i$ be the number of steps of $W'$ until vertex $i$ is first visited $\lambda/2 + 4\lambda^{3/4}$ times, and $T' = \max(T'_i) = C_n^{[A/2 + 4\lambda^{3/4}]}$. We will show that with high probability, before $T'$ steps $W'$, $W$ will have covered all vertices of $G_n$ $\lambda$ times in the required number of steps.

Let $X$ be a random variable which takes value $2j$ with probability $2^{-j}$ for $j \geq 1$, so that in particular $E(X) = 4$. Then for any $n$ and $s$, the number of steps of $W$ between steps $s - 1$ and $s$ of $W'$ is distributed as $X$. The number of steps $T$ taken by $W$ while $W'$ takes $T'$ steps is the sum of $T'$ independent copies of $X$. Thus

$$\Pr\left[T < 4(1 + 4\lambda^{-1/2})T'\right] \geq 1 - e^{-2n}.$$

Next we note that the number of hits of vertex $2i$ by $W$ per each hit of $i$ by $W'$ is distributed as $X/2$, so, by a similar argument, vertex $2i$ of $G_n$ is hit at least $\lambda + 4\lambda^{3/4}$ times, with probability at least $1 - e^{-2\sqrt{\lambda}}$, already by $T'_i$ steps of $W'$. Thus with probability at least $1 - (n/2)e^{-2\sqrt{\lambda}}$, every even vertex of $G_n$ is hit at least $\lambda + 4\lambda^{3/4}$ times.

Finally we need to cover the odd vertices of $G_n$; for this note that $W$ is equally likely to step either to $2i + 1$ or $2i - 1$ when it is at $2i$. Hence after $W$ makes $\lambda + 4\lambda^{3/4}$ visits to $2i$, with probability at least $1 - e^{-2\sqrt{\lambda}}$ that $W$ will hit vertex $2i + 1$ at least $\lambda/2$ times directly from $2i$. Thus, with probability at least $1 - ne^{-2\sqrt{\lambda}}$ each odd vertex of $G_n$ is hit at least $\lambda/2$ times from each side.

We conclude that with probability at least $1 - (3n/2 + 1)e^{-2\sqrt{\lambda}}$, $W$ covers $G_n$ in at most $4(1 + 4\lambda^{-1/2})T'$ steps. Thus

$$C_n^{[\lambda]} \leq 4(1 + 4\sqrt{\lambda})C_n^{[A/2 + 4\lambda^{3/4}]} / \left(1 - (3n/2 + 1)e^{-2\sqrt{\lambda}}\right) \leq F_n \lambda n.$$
REFERENCES


