Journal of Theoretical Probability, Vol. 2, No. 1, 1989

On the Cover Time of Random Walks on Graphs

Jeff D. Kahn,[1] Nathan Linial,[2,3] Noam Nisan,[4] and Michael E. Saks[1,5]

Received June 28, 1988

This article deals with random walks on arbitrary graphs. We consider the cover time of finite graphs. That is, we study the expected time needed for a random walk on a finite graph to visit every vertex at least once. We establish an upper bound of $O(n^2)$ for the expectation of the cover time for regular (or nearly regular) graphs. We prove a lower bound of $\Omega(n \log n)$ for the expected cover time for trees. We present examples showing all our bounds to be tight.

KEY WORDS: Random walks; cover times; graphs; infinite graphs; trees.

1. INTRODUCTION

A random walk on a graph is a very simple discrete time process. A particle starts moving on the vertices of the graph. It starts at a specific vertex and at each time step it moves from its present vertex to one of its neighbors. This neighbor is chosen at random, and all neighbors are equally likely to be selected. Such walks have been extensively studied for various highly regular graphs such as the integer line, a k-dimensional grid, Cayley graphs of various groups, etc. They are not as well understood for general graphs.

One exception is Ref. 1 dealing with random walks on general graphs. They have shown that for any undirected connected graph, a random walk will almost surely visit all the vertices in a polynomial number of steps. Specifically, they proved an upper bound of $|V|^3$ for this expectation. This

[1] Department of Mathematics, Rutgers University, New Brunswick, New Jersey.
[2] *IBM Almaden*, 650 Harry Road, San Jose, California 95120.
[3] Computer Science Department, Hebrew University, Jerusalem, Israel.
[4] Computer Science Division, University of California, Berkeley, California.
[5] Bell Communications Research, Morristown, New Jersey.
Mike Saks was supported by NSF-DMS87-03541 and by AFOSR-0271. Jeff Kahn was supported by MCS-83-01867 and by AFOSR-0271.

0894-9840/89/0100-0121$06.00/0

served as a basis for a randomized algorithm for the undirected connectivity problem that runs in logarithmic space: To determine whether two vertices in the graph are connected, the algorithm simply takes a random walk starting from the first vertex, and checks whether it arrives at the target vertex within that polynomial number of time steps.

In this paper we proceed with the investigation of the *cover time*. That is, we study the expected time needed until we first visit all the nodes in a graph. Let G be an undirected graph, and v a vertex in G. Let l be a walk originating at vertex v. We say the walk l *covers* the graph G if every vertex in G is visited at least once during the walk. For every vertex v we define a random variable X_v to be the first time a random walk originating from v *covers* the graph G.

Definition. The *cover time* of G starting from v, $c_v(G)$, is defined to be the expected value of X_v.

Reference 1 gave the first general bound for the cover time:

Theorem 1.[1] For any graph G and vertex v,

$$c_v(G) \leqslant 2\,|V|\,|E|$$

Our main results are as follows:

Theorem 2. Let $d_{\min}$ be the minimal degree in a graph G; then for any v

$$c_v(G) \leqslant 16\,\frac{|V|\,|E|}{d_{\min}}$$

Corollary 3. Let G be a regular graph, then for any v:

$$c_v(G) \leqslant O(|V|^2)$$

Theorem 3. Let G be any *tree* on n vertices, and let v be any vertex in G; then

$$c_v(G) \geqslant n \log_e n - O(n)$$

In all cases we give examples showing our results to be optimal.

2. PRELIMINARIES

Let $G=(V, E)$ be a finite graph. We discuss the following random process: A particle starts at some vertex v. At each time step it moves at random to one of the neighbors of the vertex where it currently resides. We will be interested in several properties of such a random process, most notably the expected time needed to visit all the vertices in the graph during such a process.

All graphs considered here are assumed to be connected. This causes no loss of generality, of course. We will also assume that this Markov chain is *ergodic*. This assumption simplifies the reasoning and it causes no loss of generality, since ergodicity may be guaranteed by adding a self-loop with an ε probability of stay at each vertex. This will modify the quantities we are interested in by at most a constant factor.

For any $u, v \in V$ we define the random variable $X_{u,v}$ to be the first time a random walk that originates from u reaches v. Define $T_{u,v}$ to be the expected value of $X_{u,v}$. For any $v \in V$ let $\deg(v)$ be the degree of v in G, and $N(v)$ be the set of neighbors of v in G. The following claims can be derived from standard properties of Markov chains, and are shown in Ref. 1.

Lemma 1. For every vertex $v \in V$,

$$T_{v,v} = 2\frac{|E|}{\deg(v)}$$

Corollary 1. For any vertex v,

$$\sum_{u \in N(v)} T_{u,v} = 2\,|E| - 1$$

Corollary 2. For any $(u, v) \in E$,

$$T_{u,v} \leqslant 2\,|E|$$

3. COVER TIMES FOR FINITE GRAPHS

This section deals with the following question: How long may a random walk on a finite graph take in order to visit every vertex at least once? Let l be a walk originating at vertex v. We say the walk l *covers* the graph G, if every vertex in G is visited at least once during the walk. For any vertex v we can define a random variable X_v to be the first time a random walk originating from v *covers* the graph G. We are interested in this random variable, and in particular in its expected value.

Definition. The *cover time* of G starting from v, $c_v(G)$, is defined to be the expected value of X_v.

Reference 1 gave the first general bound for the cover time:

Theorem 1.[1] For any graph G and vertex v,

$$c_v(G) \leqslant 2\,|V|\,|E|$$

The following example shows that without further constraints on the graph this result is optimal:

Example 1. Let G consist of 2 cliques of $n/3$ vertices that are connected to each other by a path of another $n/3$ vertices.

Claim. For all $v \cdot c_v(G) = \Theta(n^3)$.

The proof is left to the reader. Let us only indicate that when the walk is in one of the two cliques, the expected time to leave that clique is $\Omega(n^2)$. Also in a one-dimensional walk with a reflecting barrier at the origin the expected number of returns to the origin until first reaching k is $\Omega(k)$. The remaining easy details may be filled in by the reader.

3.1. Improved Bounds for Regular Graphs

The last example showed that the bound of $O(|V|\,|E|)$ cannot be improved without further constraints on the graph. One feature of the last example was that some of the vertices had very low degree, while others had very high degree. The following theorem shows that indeed if the degrees of the vertices are balanced then the previous bound can be improved.

Theorem 2. Let $d_{\min}$ be the minimal degree in a graph G; then for any v

$$c_v(G) \leqslant 16 \frac{|V|\,|E|}{d_{\min}}$$

Corollary 3. Let G be a regular graph then

$$c_v(G) \leqslant 8\,|V|^2$$

Using different arguments we can improve the constant in the corollary to 4. Before we prove the theorem we will need the following combinatorial lemma.

Lemma 2. Let G be a graph with minimal degree d, then there exists a collection of $d/2$ forests on V with the following properties:

1. Each forest is a subgraph of G.
2. Each edge of G appears in at most 2 forests.
3. Each forest has at most $2\,|V|/d$ components.

Proof. We construct the forests algorithmically: Start with the first component of the first forest. Pick any vertex, and let it mark one of the edges incident with it. Put this edge into the first forest and proceed to the vertex at the other end of the marked edge. Now this vertex marks one of the edges incident with it, adds this edge to the forest and we move to its other end. Continue with this procedure never marking an edge that creates a cycle. Stop this process when every edge incident with the current vertex creates a cycle, if added.

We now start the second stage of constructing the first forest. Pick any vertex that was not used in the first stage and start the same process from this vertex. It is allowed to mark an edge whose other vertex was used in the previous stage, but if this happens this stage is ended (and our forest consists at the moment of just one tree). Otherwise we mark an edge going to a new vertex and proceed with the construction. As above we stop when all edges incident with the current vertex would create a cycle if added. Now we start the next stage and so on until we use all the vertices of G. This finishes the construction of the first forest. Note that (1) this is a forest; (2) each component has at least d vertices, since all the edges incident with the last chosen vertex in a component close a cycle.

We now start constructing the second forest. It is constructed like the first one, but with the restriction that no vertex marks an edge it had marked in the construction of the previous forest. Note that an edge may appear in two forests, but each time it is marked by a different vertex. The second forest has all its components of size at least $d-1$, since all the edges going out of the last vertex in a component must close a cycle, except possibly one edge which was marked during the creation of the previous forest. We repeat this construction $d/2$ times, generating $d/2$ forests. Each component, even in the last forest, must be of size at least $d/2$, since at most $d/2-1$ of its incident edges were marked during the construction of previous forests. Thus there may be at most $2\,|V|/d$ components in each forest. Note also that each edge may appear in at most two forests since it can be marked only once by each of its two vertices. $\square$

Let us comment that this lemma may also be derived from the matroid intersection theorem (Ref. 2, p. 130).

We can prove the theorem now.

Proof. For each edge $e \in E$ where $e = (u, v)$ define the weight of e, $W(e) = T_{u,v} + T_{v,u}$. For a set of edges F, define the weight of F to be $W(F) = \sum_{e \in F} W(e)$. We will show that (1) there exists a spanning tree in G of weight at most $16 \, |V| \, |E|/d_{\min}$, and that (2) this implies a similar bound for the cover time.

We first compute the total weight of all edges in E, using Corollary 3:

$$W(E) = \sum_{e \in E} W(e) = \sum_{v \in V} \sum_{u \in N(v)} T_{u,v} = 2 \, |V| \, (|E| - 1)$$

We now invoke Lemma 2 on G, and get $d_{\min}/2$ forests: $F_1, ..., F_{d_{\min}}/2$. Since each edge appears in at most two of these forests we get that

$$\sum_i W(F_i) \leqslant 2W(E) \leqslant 4 \, |V| \, |E|$$

Thus the forest with least weight, say F, has weight bounded by $8 \, |V| \, |E|/d_{\min}$. We now wish to turn this forest into a spanning tree. This forest has at most $2 \, |V|/d_{\min}$ components. Thus it is possible to add at most $2 \, |V|/d_{\min}$ edges to it and turn it into a spanning tree, T. By Corollary 2, the weight of each such additional edge is at most $4 \, |E|$ thus the total weight added does not exceed $8 \, |V| \, |E|/d_{\min}$. Thus the total weight of T is at most $16 \, |V| \, |E|/d_{\min}$.

The cover time is upper bound by the weight of any spanning tree, as shown in Ref. 1. □

Can this result be improved? The following example shows that the result may be tight (up to a constant factor) even for d-regular graphs, of any degree d up to $n/2$.

Example 2. Let d and n be integers with $d+1$ dividing n. Consider the following d-regular graph, G_d, on n vertices: We start from $n/(d+1)$ disjoint cliques of size $d+1$ each. We now delete an edge $[a_i, b_i]$ in the ith clique. Add edges $[a_i, b_{i+1}]$, indices taken mod $n/(d+1)$.

Claim. For any d, such that $d+1$ properly divides n, and for any vertex v, $c_v(G_d) = \Theta(n^2)$.

Proof (sketch). The expected time to exit any of the cliques is clearly $\Theta(d^2)$. Our walk can be thus thought of as taking place on a circle of length $n/(d+1)$, where the probability of moving either left or right is $\Theta(d^{-2})$. Since the cover time of a circle is quadratic, the claim follows.

A trivial lower bound for the cover time in terms of $|E|/d_{\min}$ can also be given:

Proposition 1. Let $d_{\min}$ be the minimal degree in G; then there exists a vertex v such that

$$c_v(G) \geqslant \frac{2\,|E|}{d_{\min}} - 1$$

Proof. Let v be a vertex of degree $d_{\min}$; then $T_{v,v} = 2\,|E|/d_{\min}$, so the average expected time to reach v from its neighbors is $2\,|E|/d_{\min} - 1$. $\square$

3.2. Lower Bound for Trees

The previous section dealt with upper bounds for the cover time. In this section we address the question of how *small* the cover time may be. Consider the following example:

Example 3. Let G be the complete graph on n vertices, and let T be a star on n vertices, then for any v, $c_v(G) = n \log_e n + O(1)$, and $c_v(T) = 2n \log_e n + O(1)$.

Proof. Both are instances of the "coupon collector" problem.

We conjecture that $n \log n$ is indeed a lower bound for the cover time of any graph.

Conjecture 1. For any graph G, and for any vertex v,

$$c_v(G) \geqslant |V| \log_e |V| - o(|V| \log |V|)$$

We can prove this conjecture for the special case of trees.

Theorem 3. Let G be any *tree* on n vertices, and let v be any vertex in G; then

$$c_v(G) \geqslant n \log_e n - O(n)$$

Proof. Let T_n be the minimum cover time over all trees on n vertices. We will show that

$$T_n \geqslant 1 + \left(1 + \frac{1}{n-2}\right) T_{n-1}$$

The theorem follows by solving this recursion.

We first prove the theorem when v is a leaf. The general result then follows since a walk covers the tree iff it visits all the leaves. Therefore the cover time when starting at a nonleaf u is bounded below by a convex combination of the cover times starting at the leaves, the coefficient corresponding to a leaf l being the probability that it is the first leaf to be reached when starting from u.

Let x be the unique neighbor of v in G, and let G' stand for $G-v$. The cover time for G may be computed as follows: first we take one step from v to x (time taken: 1). Then we need to cover G' (expectation of the time taken: at least T_{n-1}). What needs to be added to this, in order to get the cover time of G, is the expected time spent on the edge (v, x) during the covering of G'. We will show that this time is at least $T_{n-1}/(n-2)$.

This proof is based on two facts. Let d be the degree of x in G'; then we claim:

1. The expected number of times the walk visits x before it covers G' is at least $T_{n-1}d/2(n-2)$.
2. The expected time spent taking the tour $x \to v \to x$ each time x is visited is $2/d$.

The total expected time spent on the edge (x, v) during the covering of G' will thus be the product of these two numbers, which is $T_{n-1}/(n-2)$.

Fact 1 is proven by the following argument: Define a *closed-x-tour* to be a walk in G' originating at vertex x, covering all of G', then returning to x and which is minimal with respect to these properties. Define the random variable U to be the fraction of the time spent at vertex x during a random closed-x-tour. It is clear that the expected value of U is equal to the limit probability of residing in x, which is known to be $d/2\,|E|$, where $|E|$ is the number of edges in G'. We are interested in the frequency of visiting x till G' is covered. This frequency is bounded below by U, since there are no visits to x between the time G' is covered and the first return to x. We conclude that the expected number of times x is visited before G' is covered is at least $E(U) \cdot T_{n-1} = T_{n-1}d/2(n-2)$.

The second fact is seen by observing that when x is reached, the number of times the detour $x \to v \to x$ is taken is distributed geometrically with probability $1/(d+1)$. The cost of each detour is 2, so the expected cost is $2/d$. □

REFERENCES

1. Aleliunas, R., Karp, R. M., Lipton, R. J., Lovasz, L., and Rackoff, C. (1979). Random walks, universal traversal sequences, and the complexity of maze problems, 20'th FOCS.
2. Welsh, D. J. A. (1976). *Matroid Theory*. Academic Press, London.