MAX-PLANCK-INSTITUT FÜR INFORMATIK

Short Random Walks on Graphs

G. Barnes U. Feige

MPI-I-94-121 April 1994

INFORMATIK

Im Stadtwald
66123 Saarbrücken
Germany

Short Random Walks on Graphs

G. Barnes U. Feige

MPI–I–94–121 April 1994

Short Random Walks on Graphs*

Greg Barnes† Uriel Feige‡

March 15, 1994

Abstract

We study the short term behavior of random walks on graphs, in particular, the rate at which a random walk discovers new vertices and edges. We prove a conjecture by Linial that the expected time to find $\mathcal{N}$ distinct vertices is $O(\mathcal{N}^3)$. We also prove an upper bound of $O(\mathcal{M}^2)$ on the expected time to traverse $\mathcal{M}$ edges, and $O(\mathcal{M}\mathcal{N})$ on the expected time to either visit $\mathcal{N}$ vertices or traverse $\mathcal{M}$ edges (whichever comes first).

1 Introduction

Consider a simple random walk on G, an undirected graph with n vertices and m edges. At each time step, if the walk is at vertex v, it moves to a vertex chosen uniformly at random from the neighbors of v. Random walks have been studied extensively, and have numerous applications in theoretical computer science, including space-efficient algorithms for undirected connectivity [4, 8], derandomization [1], recycling of random bits [10, 15], approximation algorithms [6, 12, 17], efficient constructions in cryptography [14], and self-stabilizing distributed computing [11, 16].

Frequently (see, for example, Karger *et al.* [19] and Nisan *et al.* [20]), we are interested in $E[T(\mathcal{N})]$, the expected time before a simple random walk on an undirected connected graph, G, visits its $\mathcal{N}^{th}$ distinct vertex, $\mathcal{N} \leq n$. The corresponding question for edges is also interesting, and arises in the work of Broder *et al.* [8]: how large is $E[T(\mathcal{M})]$, the expected time before a simple random walk on an undirected connected graph, G, traverses its $\mathcal{M}^{th}$ distinct edge, $\mathcal{M} \leq m$? This paper gives upper bounds on $E[T(\mathcal{N})]$ and $E[T(\mathcal{M})]$ for arbitrary graphs. While a great deal was previously known about how quickly a random walk covers the entire graph (see, for example, [2, 4, 7, 9, 18, 22, 23]), little was known

*Portions of this work were performed while the authors were at the IBM T.J. Watson Research Center, Yorktown Heights, NY.

†Max-Planck-Institut für Informatik, Im Stadtwald, 66123 Saarbrücken, Germany. Email address: barnes@mpi-sb.mpg.de.

‡Department of Applied Mathematics, The Weizmann Institute, Rehovot 76100, Israel. Email address: feige@wisdom.weizmann.ac.il. Supported by a Koret Foundation fellowship.

about the behavior of a random walk before the vertices are covered. These bounds help fill the gaps in our knowledge of random walks, giving a picture of the rate at which a random walk explores a finite or an infinite graph.

Aleliunas *et al.* [4] show that the expected time to visit *all* vertices of an arbitrary graph (called the *cover time*) is $O(mn) \leq O(n^3)$. Using this bound, Linial derives a bound for general $\mathcal{N}$ of $E[T(\mathcal{N})] = O(\mathcal{N}^4)$ [19, Lemma 4.1]. Linial [personal communication] conjectures that the cover time bound generalizes to all $\mathcal{N}$, that is, $\forall \mathcal{N} \leq n, E[T(\mathcal{N})] = O(\mathcal{N}^3)$. We prove Linial's conjecture.

Theorem 1 *For any connected graph on n vertices, and for any $\mathcal{N} \leq n$,*

$$E[T(\mathcal{N})] = O(\mathcal{N}^3).$$

Zuckerman [23] proves an upper bound of $O(mn)$ on the time to traverse all edges in a general graph. We are unaware of any previous nontrivial bounds for $\mathcal{M} < m$. We prove:

Theorem 2 *For any connected graph with m edges, and for any $\mathcal{M} \leq m$,*

$$E[T(\mathcal{M})] = O(\mathcal{M}^2).$$

Theorem 2 holds even if G is not a simple graph (i.e., if we allow self-loops and parallel edges).

Let $E[T(\mathcal{M},\mathcal{N})]$ be the expected time for a simple random walk to either traverse $\mathcal{M}$ distinct edges or visit $\mathcal{N}$ distinct vertices (whichever comes first). Then the following theorem implies both the above theorems, by considering $E[T(\mathcal{N}^2,\mathcal{N})]$ and $E[T(\mathcal{M},\mathcal{M})]$, respectively.

Theorem 3 *For any connected graph with m edges and n vertices, and for any $\mathcal{M}$ and $\mathcal{N}$ such that $\mathcal{M} \leq m$ or $\mathcal{N} \leq n$,*

$$E[T(\mathcal{M},\mathcal{N})] = O(\mathcal{M}\mathcal{N}).$$

In the above three theorems, the graph G need not be finite. If G is a graph with infinitely many vertices (each vertex of finite degree), then we can consider only the finite portion of G that is within distance $\mathcal{N}$ (or $\mathcal{M}$) from the starting vertex of the random walk, and the proofs remain unchanged. For finite graphs, the following theorem serves to complete the picture of the rate at which vertices (or edges) are discovered. It provides better bounds than Theorems 1 and 2 when the number of vertices to be discovered is larger than $\sqrt{m}$ or the number of edges to be discovered is larger than n.

Theorem 4 *For any simple connected graph on n vertices and m edges, for any $\mathcal{N} \leq n$,*

$$E[T(\mathcal{N})] = O(m\mathcal{N}),$$

and for any $\mathcal{M} \leq m$,

$$E[T(\mathcal{M})] = O(n\mathcal{M}).$$

Our theorems are the best possible in the sense that there exist graphs for which the bounds are tight up to constant factors (e.g., the n-cycle for Theorem 2). However, these bounds can be refined if additional information regarding the structure of G is given. The work of Kahn *et al.* [18] indicates that d_{min}, the minimum degree of the vertices in the graph G, is a useful parameter to consider. They show that the expected cover time of any connected graph is $O(mn/d_{min})$, implying a cover time of $O(n^2)$ for regular graphs. This inverse dependency on d_{min} applies also to short random walks. Preliminary results in this direction (tight up to a logarithmic factor) were presented in an earlier version of this paper [5]. The superfluous logarithmic factor in these results was subsequently removed by Feige [13], building upon proof techniques that were developed by Aldous [3]. Aldous is writing a textbook giving a systematic account of random walks on graphs and reversible Markov chains. The current draft [3] contains results similar to ours in the regular graph setting.

While the short term behavior of random walks is worth studying in its own right, short random walks also have immediate applications in many areas of computer science. Our results, of course, cannot be applied to all such areas. For example, much stronger results are already known about the properties of short random walks on the special class of graphs known as *expanders* (see, for example, Ajtai *et al.* [1], and Jerrum and Sinclair [17]). One might hope our results would dramatically improve the algorithms of Karger *et al.* [19] and Nisan *et al.* [20] for undirected connectivity. As mentioned above, both require an estimate of $E[T(\mathcal{N})]$ (and both used the estimate $E[T(\mathcal{N})] = O(n^4)$). Unfortunately, substituting our bound only improves the constants for the algorithms, since the running times of both depend on the *logarithm* of $E[T(\mathcal{N})]$, not $E[T(\mathcal{N})]$.

Our results may yield significant improvements for other randomized algorithms. In particular, consider randomized time-space tradeoffs for undirected $\mathcal{S}$-$\mathcal{T}$ connectivity (UST-CON), as studied by Broder *et al.* [8]. One key property of Broder *et al.*'s algorithm is that a short random walk from a given edge traverses many edges. Improved bounds on $E[T(\mathcal{M})]$, then, would seem to provide an improvement to their tradeoff. Partial results in this direction were presented in an earlier version of our paper [5], and further improvements are presented by Feige [13].

2 Short random walks - proofs of theorems

The proofs of the theorems are best read in order. The proof of Theorem 1 introduces the proof techniques that are used in all subsequent proofs. The proof of Theorem 2 is a simple modification of this proof technique. The proof of Theorem 3 introduces additional subtleties. For this reason, we have stated Theorems 1 and 2 explicitly, rather than presenting them as corollaries of Theorem 3. The proof of Theorem 4 is a slight modification of the proofs of Theorems 1 and 3.

Proof:(of Theorem 1)

Assume that $n > 2\mathcal{N}$. Otherwise the proof follows from Aleliunas *et al.*'s bound of $E[T(n)] = O(mn)$ [4]. We view the random walk as proceeding in phases. For any i, $1 \leq i \leq 2\mathcal{N}$, at the beginning of Phase i, we identify a set of vertices $V_i \subset V$ and a starting vertex $s_i \in V_i$, the last vertex visited in Phase $i-1$. Phase i starts with the random walk at s_i and ends when the random walk exits V_i. We show that for any i, the expected number of steps taken in Phase i is $O(i^2)$, and that up to Phase i at least $i/2$ distinct vertices are visited. Thus, at most $2\mathcal{N}$ phases are needed to visit $\mathcal{N}$ distinct vertices, proving (by linearity of the expectation) that $E[T(\mathcal{N})] = O(\mathcal{N}^3)$.

To simplify the presentation, assume that G contains a Hamiltonian cycle. The case where G does not have a Hamiltonian cycle is only slightly more complicated, and will be addressed later. Let v_1, v_2, ..., v_n be an arrangement of the vertices of G along the Hamiltonian cycle.

At the beginning of Phase i, we identify the following vertices and sets of vertices.

starting vertex The vertex s_i at which the walk of Phase $i-1$ ended (if $i > 1$). s_1 is the starting vertex of the whole random walk.

right vertex The vertex r_i following s_i in the cyclic order imposed by the Hamiltonian cycle.

visited vertices The set $Y_i \subset V$ of vertices visited in previous steps by the walk. Note that $s_i \in Y_i$.

good vertices The set of vertices $U_i \subseteq V \setminus Y_i$ (for two sets A and B, $A \setminus B$ denotes the set of elements in A but not in B) with the following property: let $R_{v_j,\ell}$, for $1 \leq \ell \leq n$, be the set of ℓ consecutive vertices $\{v_j, v_{j+1}, \ldots, v_{j+\ell-1}\}$ on the Hamiltonian cycle. (By convention, if $k > n$, then v_k is interpreted as v_{k-n}.) $v_j \in U_i$ if and only if $\forall \ell \leq n, |R_{v_j,\ell} \bigcap Y_i| \leq \ell/2$. Thus, a vertex v_j is good in Phase i if, starting at v_j and walking along the Hamiltonian cycle, at least half of the vertices discovered are new. This holds for any number of steps that a walk might make before walking completely around the cycle.

bad vertices All other vertices. Let B_i denote this set.

The following lemma gives a bound on $|B_i|$.

Lemma 5 *For every i, if $|Y_i| < 2n$, then $|B_i| < |Y_i|$.*

Proof: The proof is by induction on $|Y_i|$. If $|Y_i| = 1$ (and $n > 2$), then $|B_i| = 0$, as there is no consecutive set of vertices starting from an unvisited vertex such that the majority of vertices in the set are not visited.

Assume the lemma is true for $|Y_i| = k$ and any $n > 2k$, and prove for $|Y_i| = k + 1$ and any $n > 2k + 2$. Consider a ring with n vertices and $k + 1$ visited vertices ($n > 2k + 2$). Then there exists a visited vertex y that is to the right of an unvisited vertex v. Remove y and v from the ring, linking the left neighbor of v to the right neighbor of y in the obvious way. Thus, n is decreased by 2 and $|Y_i|$ is decreased by 1. Now the induction hypothesis holds, and there are at most $k - 1$ bad vertices. Put y and v back in the ring, and restore the values of $|Y_i|$ and n. No previously good vertex can become bad, as any consecutive set of vertices that starts at an unvisited vertex and includes y must contain v as well. Thus, the only bad vertex that could possibly have been added is v, resulting in the desired bound of at most k bad vertices. □

Let $V_i = (Y_i \bigcup B_i) \setminus \{r_i\}$. At the beginning of Phase i the random walk is at the vertex $s_i \in V_i$. Phase i ends when the random walk exits V_i. Let T_i denote the number of steps taken in Phase i.

Lemma 6 $E[T_i] < (2|Y_i|)^2$.

Proof: By Lemma 5, and since $|V_i| \leq |Y_i| + |B_i|$, it follows that $|V_i| < 2|Y_i|$. Phase i ends when an edge leading out of V_i is taken. One such edge is the edge connecting s_i to r_i on the Hamiltonian cycle. We claim that if we remove all other edges leading out of V_i, the expected time to leave V_i remains at least $E[T_i]$.

Suppose we wish to remove a single edge e between $u \in V_i$ and $v \notin V_i$. Instead of actually removing e, create a new auxiliary vertex $w \notin V_i$ and make e connect u with w. This does not change $E[T_i]$. Let T' denote the number of steps taken to exit $V_i \bigcup \{w\}$, not counting steps that traverse the edge e (in either direction). Then $E[T'] \geq E[T_i]$. Finally, remove e completely, and observe that the expected time to leave V_i remains $E[T']$.

After removing all edges but $e = \{s_i, r_i\}$ leading out of V_i, we are left with only the subgraph induced by V_i and the edge e. Let m_i denote the number of edges in this subgraph. Observe that if $i > 1$, $m_i < |V_i|^2/2$. (If $i = 1, T_i = 1$ and the lemma trivially holds.) Since the walk of Phase i starts at s_i, and since there is an edge connecting s_i to r_i, the expected time to reach r_i is at most $2m_i$. (This is well known. See, for example, Aleliunas *et al.* [4].) □

We have bounded the duration of each phase. It remains to bound the number of phases.

Lemma 7 *For any k, in the first k phases at least $k/2$ distinct vertices are visited.*

Proof: Observe that each phase ends by either walking to the next vertex along the Hamiltonian cycle, or "jumping" to a good vertex. Let S be the sequence of starting vertices visited by the walk, $s_1, s_2, \ldots, s_k$. Partition S into subsequences as follows: the first subsequence begins with s_1, and s_j begins a new subsequence if and only if Phase $j-1$ did not end at r_{j-1}. Then each subsequence begins with a good vertex, and at least half the vertices visited in it are new. □

If G has a Hamiltonian cycle, this completes the proof, as

$$E[T(\mathcal{N})] \leq \sum_{i=1}^{2\mathcal{N}} E[T_i] < \sum_{i=1}^{2\mathcal{N}} (2|Y_i|)^2 < 8\mathcal{N}^3.$$

If G does not have a Hamiltonian cycle, we can use the following well-known lemma:

Lemma 8 *For any connected graph, G, there is a cyclic ordering of its vertices, w_1, w_2, ..., w_n, such that the distance (the length of the shortest path in G) between any vertex and its successor is at most 3.*

Proof: Let G_{span} be a spanning tree of G. Traverse G_{span} in depth-first search fashion, using vertices of even distance from the root to advance towards the leaves, and vertices of odd distance from the root to backtrack. Let $w_1, w_2, \ldots, w_n$ be the vertex ordering derived from this traversal, where w_i is the i^{th} vertex visited by the traversal. Then w_n is a neighbor of w_1, the root of G_{span}, and for all $1 \leq i < n$, w_i is at most distance 3 from w_{i+1}. □

Using the ring obtained by Lemma 8 in place of a Hamiltonian cycle makes the expected time to leave V_i at most three times as large, thus only affecting the constants involved. □

Proof: (**of Theorem 2**)

In Lemma 8 we show a way to arrange the vertices of a connected graph in a cyclic order. Arranging the edges in a cyclic order is even simpler. View each undirected edge as two *anti-parallel* directed edges (two directed edges are anti-parallel if they have the same endpoints, u and v, but one is directed from u to v, and the other directed from v to u). The number of directed edges entering any vertex is equal to the number of directed edges leaving it. Hence the directed graph is Eulerian, and has an Eulerian cycle. This Eulerian cycle induces a cyclic ordering on the directed edges, and can replace the Hamiltonian cycle used in the proof of Theorem 1.

Now the proof technique of Theorem 1 can be applied to prove Theorem 2, with "directed edges" replacing "vertices" in a straightforward manner. For edges, however, Lemma 6 can be strengthened — in Phase i, the set V_i is now a set of edges, and not a set of vertices, so the expected time to leave V_i is $|V_i|$ instead of $|V_i|^2$. This yields a bound of $E[T(\mathcal{M})] = O(\mathcal{M}^2)$. □

Proof:(of Theorem 3)
Assume that $m \geq 2\mathcal{M}$. The proof for the case $m < 2\mathcal{M}$ is much simpler. See, for example, the first part of the proof of Theorem 4, below.

As in the proof of Theorem 1, view the random walk as proceeding in phases. For any $i \geq 1$, at the beginning of Phase i, we identify a set of vertices $V_i \subset V$, a set of edges $E_i \subset E$ and a starting vertex $s_i \in V_i$, the last vertex visited in Phase $i-1$. Phase i starts with the random walk at s_i. Phase i ends when the random walk exits the subgraph $G_i = (V_i, E_i)$. Phases where s_i has many yet unvisited outgoing edges will also end if the walk returns to s_i. The whole walk ends when either $\mathcal{N}$ vertices or $\mathcal{M}$ edges are visited. This does not necessarily correspond to any fixed number of phases, making the analysis of the expected number of steps taken by the walk more subtle than the analysis in the proof of Theorem 1, where the completion of $2\mathcal{N}$ phases guaranteed the end of the walk.

View each undirected edge as two anti-parallel directed edges. Observe that traversing $2\mathcal{M}-1$ distinct directed edges guarantees at least $\mathcal{M}$ distinct original (undirected) edges are traversed. The set of outgoing edges from vertex v is denoted by $Out(v)$. Let $d(v) = |Out(v)|$.

Let $v_1, v_2, \ldots, v_n$ be an arrangement of the vertices of G along the ring obtained by Lemma 8. At the beginning of Phase i, we identify the following vertices and sets of vertices.

starting vertex The vertex s_i at which the walk of Phase $i-1$ ended (if $i > 1$). s_1 is the starting vertex of the whole random walk.

right vertex The vertex r_i following s_i in the ring of vertices.

traversed edges The set $F_i \subset E$ of edges traversed in previous steps by the walk.

visited vertices The set $Y_i \subseteq V$ of vertices visited in previous steps by the walk. Note that $s_i \in Y_i$.

exhausted vertices The set $X_i \subseteq Y_i$ of vertices, x, such that at least half of the out-edges from x have been traversed in previous steps by the walk.

good vertices The set of vertices U_i with the following property: as in the proof of Theorem 1, let $R_{v_j,\ell}$, for $1 \leq \ell \leq n$, be the set of ℓ consecutive vertices $\{v_j, v_{j+1}, \ldots, v_{j+\ell-1}\}$ on the ring of vertices. Define $g_i(v)$ to be $max(0, \lceil d(v)/2 \rceil - |F_i \bigcap Out(v)|)$, that is, the number of untraversed out-edges of v that would have to be traversed for v to become exhausted. For a vertex v_j, $v_j \in U_i$ if and only if $\forall \ell, 1 \leq \ell \leq n$, $\sum_{v_k \in R_{v_j,\ell}} g_i(v_k) \geq (2\mathcal{M}/\mathcal{N})|R_{v_j,\ell} \bigcap X_i|$.

Note that a visited vertex can be good, but an exhausted vertex cannot. Note also that s_1 is a good vertex in Phase 1.

bad vertices The set of vertices B_i that are not good.

good edges The set of untraversed edges $D_i \subseteq E \setminus F_i$ that exit from good vertices.

bad edges The set of edges C_i that are neither good nor traversed.

Informally, the definition of good vertices above is similar to the definition of good vertices in the proof of Theorem 1. In that proof, each subsequence (see the proof of Lemma 7) begins at a good vertex, and progresses along the ring. The definition of a good vertex insures that the number of previously unvisited vertices visited during the subsequence is at least a constant fraction of the number of phases in the subsequence. In this proof, a subsequence again starts at a good vertex, but the walk does not progress along the ring until the current vertex is exhausted. This definition of good vertices ensures that the number of previously untraversed edges that are traversed during a subsequence is at least $2\mathcal{M}/\mathcal{N}$ times the number of exhausted vertices that are starting vertices in the phases in the subsequence. This property is used in Lemma 13 below to bound the number of phases that begin at exhausted vertices.

The following lemma gives a bound on $|C_i|$.

Lemma 9 *For every i, $|C_i| \leq |F_i| + 4\mathcal{M}(1 + |X_i|/\mathcal{N})$. In particular, for $|F_i| < 2\mathcal{M}$, $|C_i| < 10\mathcal{M}$.*

Proof: Consider the ring of vertices at the beginning of Phase i. For any vertex v, let $g_i'(v) = max(0, d(v) - 2|F_i \bigcap Out(v)|)$, and mark $g_i'(v)$ of v's untraversed out-edges. Since $2|F_i| + \sum_v g_i'(v) \geq 2m$, the number of untraversed edges in G that are not marked is at most $|F_i|$. We will show by induction on $|X_i|$ that the number of marked edges that are bad is no more than $4\mathcal{M}(1 + |X_i|/\mathcal{N})$. Hence the total number of bad edges is as claimed in the lemma.

To bound the number of marked edges that are bad, we prove the following lemma. The lemma is more general than is necessary, since it considers not only configurations of marked edges and exhausted vertices that could be a result of random walks on graphs, but also configurations that could not.

Lemma 10 *Consider a ring of n vertices, and let $k < \mathcal{N} \leq n$. Choose k of the vertices in an arbitrary way and mark them as* exhausted. *Distribute an arbitrary number of tokens on the unexhausted vertices of the ring in an arbitrary way. A vertex v is* bad *if for some $\ell \leq n$, the number of tokens encountered by taking ℓ steps to the right (including the tokens on v itself) is less than $4\mathcal{M}/\mathcal{N}$ times the number of exhausted vertices encountered by such* *a walk. A token is* bad *if it is placed on a bad vertex. Then for any $n, \mathcal{N}, \mathcal{M}, k$, the number of bad tokens is at most $4\mathcal{M}(1 + k/\mathcal{N})$.*

Proof: By induction on k. If $k = 0$, then for all values of n, $\mathcal{N}$ and $\mathcal{M}$, there are no bad tokens, and the lemma holds.

Assume the lemma is true for $k = j$, for all values of n, $\mathcal{N}$ and $\mathcal{M}$, where $k < \mathcal{N} \leq n$. Prove for $k = j + 1$ and arbitrary n, $\mathcal{N}$ and $\mathcal{M}$, where $k < \mathcal{N} \leq n$.

Consider a walk backwards along the ring from some exhausted vertex, v. After a certain number of steps along the ring, the walk will have encountered $4\mathcal{M}/\mathcal{N}$ tokens. Let y_v be the vertex where this walk from v first reaches its $(4\mathcal{M}/\mathcal{N})^{th}$ token. Because there are at least $4\mathcal{M}$ tokens (otherwise, the lemma trivially holds) and at most $\mathcal{N}$ exhausted vertices, there must be some exhausted vertex v such that the walk from v to y_v visits no exhausted vertex besides v.

Let v be such a vertex, and $\mathcal{T}_v$ be the first $4\mathcal{M}/\mathcal{N}$ tokens encountered by this walk (this may include only some of the tokens placed on y_v). Remove from the ring the tokens $\mathcal{T}_v$, and make v not exhausted. Thus, k is decreased by 1. Now the induction hypothesis holds, so there are at most $4\mathcal{M}(1 + (k - 1)/\mathcal{N})$ bad tokens.

Add the tokens in $\mathcal{T}_v$ back to the ring, mark v as exhausted, and restore the value of k. The tokens in $\mathcal{T}_v$ may be bad, but no token t that is not in $\mathcal{T}_v$ and was not previously bad can become bad, as any walk from t that includes v must include all the tokens between y_v and v as well. So the number of bad tokens increases by at most $4\mathcal{M}/\mathcal{N}$, proving the lemma. □

Observe that for any subset V_s of the vertices, $\sum_{v \in V_s} g_i(v) \geq (\sum_{v \in V_s} g_i'(v))/2$. By considering marked edges as the tokens of Lemma 10 and bad marked edges as the bad tokens, the proof of Lemma 9 follows. □

The definition of the subgraph $G_i = (V_i, E_i)$ and the stopping condition for Phase i depends on whether s_i is exhausted or not. At the beginning of Phase i, the random walk is at the vertex $s_i \in V_i$.

If s_i is not exhausted, $V_i = B_i \bigcup \{s_i\}$, E_i is all edges with both endpoints in V_i, along with the edges out of s_i and the edges into s_i, and Phase i ends when the random walk returns to s_i or exits G_i by visiting a vertex in U_i.

If s_i is exhausted, $V_i = B_i \setminus \{r_i\}$, E_i is all edges with both endpoints in V_i, along with the edges along a shortest path from s_i to r_i and the edges anti-parallel to the edges in this path, and Phase i ends when the random walk exits G_i by visiting a vertex in $U_i \bigcup \{r_i\}$.

Let T_i denote the number of steps taken in Phase i.

Lemma 11 *If Phase i begins at an unexhausted vertex, v, $E[T_i] < 12\mathcal{M}/d(v) + 2$.*

Proof: The number of edges in G_i is no more than the number of out-edges from vertices in V_i, plus the edges into s_i. An out-edge from a bad vertex is either bad or traversed, so $|E_i| \leq |F_i| + |C_i| + 2d(v)$. Therefore, by Lemma 9, if $|F_i| < 2\mathcal{M}, |E_i| < 12\mathcal{M} + 2d(v)$.

It is well known that on an undirected graph with m_i edges the expected time for a random walk that starts at vertex v to return to v is $2m_i/d(v)$. G_i is equivalent to an undirected graph with $|E_i|/2$ edges, since if a directed edge e is in E_i, the edge anti-parallel to e is in E_i as well. The degree of v in G_i is the same as its degree in G, therefore the expected length of a phase that begins at an unexhausted vertex, v, is no more than $12\mathcal{M}/d(v) + 2$. □

Lemma 12 *If Phase i begins at an exhausted vertex,* $E[T_i] < 36\mathcal{M} + 15$.

Proof: The only edges in E_i that may not out-edges from bad or exhausted vertices are the edges in the path from s_i to r_i and the edges anti-parallel to these edges. By the construction of the ring of vertices, this path is of length 3 or less, and the first out-edge in the path from s_i is from an exhausted vertex, so there are at most five such edges, and $|E_i| \leq |F_i| + |C_i| + 5 < 12\mathcal{M} + 5$. Using a proof similar to the proof of Lemma 6, the expected length of such a phase is less than the distance from s_i to r_i times $|E_i|$. □

The expected number of steps in Phase i depends on whether s_i is exhausted or not. Call phases that start at unexhausted vertices *short phases* and phases that start at exhausted vertices *long phases.*

The walk ends when $|Y_i| \geq \mathcal{N}$ or $|F_i| \geq 2\mathcal{M} - 1$. Since we are considering directed edges, this will ensure that either $\mathcal{N}$ vertices were visited or $\mathcal{M}$ undirected edges were traversed. To analyze the expected number of steps of the walk, we consider the following two stopping conditions:

1. At least $\mathcal{N}$ distinct vertices were the starting vertices of phases. Clearly, this implies that $\mathcal{N}$ vertices were visited.
2. There were at least $2\mathcal{N}$ long phases. This implies that at least $2\mathcal{M}$ edges have been visited, by the following lemma.

Lemma 13 *If $m \geq 2\mathcal{M}$, $2\mathcal{N}$ long phases occur, and no more than $\mathcal{N}$ distinct vertices are the starting vertices of these long phases, then at least $2\mathcal{M}$ edges are traversed.*

Proof: Similar to the proof of Lemma 7, observe that each phase ends by either walking to the next vertex in the ring, returning to the starting vertex, or "jumping" to a good vertex. Let S be the sequence of starting vertices visited by the walk, $s_1, s_2, \ldots, s_k$. Partition S into subsequences as follows: the first subsequence begins with s_1, and s_j begins a new subsequence if and only if Phase $j-1$ was long and did not end at r_{j-1}, or Phase $j-1$ was short and did not end at s_{j-1}. Then each subsequence begins with a good vertex (and a short phase), and continually exhausts its current vertex and steps to the next vertex in the ring. Because $m \geq 2\mathcal{M}$, a subsequence that begins at s_j cannot step completely around the ring and return to s_j. We use the following property: For a subsequence S_j beginning with Phase k, at least $2q(\mathcal{M}/\mathcal{N})$ edges are traversed for the first time in the phases of S_j before the q^{th} vertex in X_k appears as a starting vertex in S_j. This must be true by the definition of a good vertex. With each vertex v of X_k we can therefore associate $2\mathcal{M}/\mathcal{N}$ untraversed edges that must be traversed if v is to start a phase within the subsequence, associating each untraversed edge with at most one vertex of X_k.

Now consider the $2\mathcal{N}$ long phases. They start at no more than $\mathcal{N}$ distinct vertices, so at least $\mathcal{N}$ of them start at a vertex that was the starting vertex of a previous long phase. Let Phase i be such a phase, and assume that it is part of a subsequence that begins with Phase k. Then s_i must already have been exhausted when Phase k begun, implying that we can identify $2\mathcal{M}/\mathcal{N}$ edges that were first traversed between Phases k and i, and associate them with Phase i. Altogether, from all long phases, we can identify at least $\mathcal{N} \cdot 2\mathcal{M}/\mathcal{N} = 2\mathcal{M}$ distinct traversed edges. □

We are now ready to compute the expected number of steps of the walk. There are at most $2\mathcal{N}$ long phases, and each long phase takes expected time no more than $36\mathcal{M} + 15$, so the long phases contribute a total of $O(\mathcal{N}\mathcal{M})$ to the expected number of steps in the walk. To analyze the contribution of the short phases, let v_i denote the i^{th} distinct vertex that was discovered by the walk, and let $E[v_i]$ denote the expected number of short phases that start at v_i. For each such phase, the probability that the first step of the walk traverses a yet untraversed out-edge from v_i is at least $1/2$, since the majority of edges leading out of v_i are untraversed. Therefore $E[v_i]$ is no more than $d(v_i)$. If $d(v_i) > 2\mathcal{M}$, similar reasoning shows that $E[v_i]$ is no more than $2\mathcal{M}$, since the walk can stop after $\mathcal{M}$ distinct out-edges of v_i are traversed. The expected number of steps in a short phase is no more than $12\mathcal{M}/d(v_i) + 2$, so it follows by Wald's Equation (see Ross [21, page 38], for example) that the expected number of steps spent on short phases that begin at v_i is no more than $(12\mathcal{M}/d(v_i) + 2) \cdot \min(d(v_i), 2\mathcal{M}) \leq 16\mathcal{M}$. The short phases therefore contribute a total of $O(\mathcal{N}\mathcal{M})$ to the expected number of steps in the walk. □

Proof: (of Theorem 4)

To show that $E[T(\mathcal{N})] = O(m\mathcal{N})$, we distinguish between two cases. The case $\mathcal{N} \geq n/2$ is handled by Aleliunas *et al.* [4], who show that $E[T(n)] = O(mn)$. For the case $\mathcal{N} < n/2$, consider the proof of Theorem 1. The expected time of each phase is at most $6m$, and the proof follows.

To show that $E[T(\mathcal{M})] = O(n\mathcal{M})$, we again distinguish between two cases. The case $\mathcal{M} > m/2$ is handled by Zuckerman [23], who shows that $E[T(m)] = O(mn)$. For the case $\mathcal{M} \leq m/2$, consider the proof of Theorem 3 with $\mathcal{N} = n$. Observe that by Lemma 13, in order to visit $\mathcal{M}$ distinct edges, it suffices to have $2n$ phases that start at exhausted vertices. Hence the expected number of steps spent on phases that start at exhausted vertices is $O(\mathcal{M}n)$. Likewise, since the graph has only n vertices, the expected number of steps spent on phases that start at vertices that are not exhausted is also $O(\mathcal{M}n)$ (the fact that the random walk may not stop at the time that all vertices of G are visited does not affect this argument). The proof follows. □

Acknowledgements

We thank Nati Linial, Prabhakar Raghavan and Larry Ruzzo for many helpful discussions, and for directing us to relevant literature.

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi. Deterministic simulation in LOGSPACE. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, pages 132–140, New York, NY, May 1987.

[2] D. J. Aldous. Lower bounds for covering times for reversible Markov chains and random walks on graphs. *Journal of Theoretical Probability*, 2(1):91–100, Jan. 1989.

[3] D. J. Aldous. *Reversible Markov chains and random walks on graphs.* 1993. Draft, University of California, Berkeley.

[4] R. Aleliunas, R. M. Karp, R. J. Lipton, L. Lovász, and C. Rackoff. Random walks, universal traversal sequences, and the complexity of maze problems. In *20th Annual Symposium on Foundations of Computer Science*, pages 218–223, San Juan, Puerto Rico, Oct. 1979. IEEE.

[5] G. Barnes and U. Feige. Short random walks on graphs. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing*, pages 728–737, San Diego, CA, May 1993.

[6] A. Z. Broder. How hard is it to marry at random? (On the approximation of the permanent). In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, pages 50–58, Berkeley, CA, May 1986. See also *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing,* page 551, Chicago, IL, May 1988.

[7] A. Z. Broder and A. R. Karlin. Bounds on the cover time. *Journal of Theoretical Probability*, 2(1):101–120, Jan. 1989.

[8] A. Z. Broder, A. R. Karlin, P. Raghavan, and E. Upfal. Trading space for time in undirected s-t connectivity. In *Proceedings of the Twenty First Annual ACM Symposium on Theory of Computing*, pages 543–549, Seattle, WA, May 1989.

[9] A. K. Chandra, P. Raghavan, W. L. Ruzzo, R. Smolensky, and P. Tiwari. The electrical resistance of a graph captures its commute and cover times. In *Proceedings of the Twenty First Annual ACM Symposium on Theory of Computing*, pages 574–586, Seattle, WA, May 1989.

[10] A. Cole and A. Wigderson. Dispersers, deterministic amplification, and weak random sources. In *30th Annual Symposium on Foundations of Computer Science*, pages 14–19, Research Triangle Park, NC, Oct. 1989. IEEE.

[11] D. Coppersmith, P. Tetali, and P. Winkler. Collisions among random walks on a graph. Technical report, DIMACS, 1992.

[12] M. Dyer, A. Frieze, and R. Kannan. A random polynomial time algorithm for approximating the volume of convex bodies. In *Proceedings of the Twenty First Annual ACM Symposium on Theory of Computing*, pages 375–381, Seattle, WA, May 1989.

[13] U. Feige. A randomized time-space tradeoff of $\tilde{O}(m\hat{R})$ for USTCON. In *34th Annual Symposium on Foundations of Computer Science*, pages 238–246, Palo Alto, CA, Nov. 1993. IEEE.

[14] O. Goldreich, R. Impagliazzo, L. Levin, R. Venkatesan, and D. Zuckerman. Security preserving amplification of hardness. In *31st Annual Symposium on Foundations of Computer Science*, pages 318–326, St. Louis, MO, Oct. 1990. IEEE.

[15] R. Impagliazzo and D. Zuckerman. How to recycle random bits. In *30th Annual Symposium on Foundations of Computer Science*, pages 248–253, Research Triangle Park, NC, Oct. 1989. IEEE.

[16] A. Israeli and M. Jalfon. Token management schemes and random walks yield self-stabilizing mutual exclusion. In *Proceedings of the Ninth Annual ACM Symposium on Principles of Distributed Computing*, pages 119–131, 1990.

[17] M. Jerrum and A. Sinclair. Conductance and the rapid mixing property for Markov chains: The approximation of the permanent resolved. In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, pages 235–244, Chicago, IL, May 1988.

[18] J. D. Kahn, N. Linial, N. Nisan, and M. E. Saks. On the cover time of random walks on graphs. *Journal of Theoretical Probability*, 2(1):121–128, Jan. 1989.

[19] D. R. Karger, N. Nisan, and M. Parnas. Fast connected components algorithms for the EREW PRAM. In *Proceedings of the 1992 ACM Symposium on Parallel Algorithms and Architectures*, pages 373–381, San Diego, CA, June–July 1992.

[20] N. Nisan, E. Szemerédi, and A. Wigderson. Undirected connectivity in $O(\log^{1.5} n)$ space. In *33rd Annual Symposium on Foundations of Computer Science*, pages 24–29, Pittsburgh, PA, Oct. 1992. IEEE.

[21] S. M. Ross. *Applied Probability Models with Optimization Applications.* Holden-Day, San Francisco, CA, 1970.

[22] D. I. Zuckerman. A technique for lower bounding the cover time. In *Proceedings of the Twenty Second Annual ACM Symposium on Theory of Computing*, pages 254–259, Baltimore, MD, May 1990.

[23] D. Zuckerman. On the time to traverse all edges of a graph. *Information Processing Letters*, 38(6):335–337, 28 June 1991.