Random Walks on Colored Graphs

Anne Condon * Diane Hernek †

March 1, 1994

Abstract

We initiate a study of random walks on undirected graphs with colored edges. In our model, a sequence of colors is specified before the walk begins, and it dictates the color of edge to be followed at each step. We give tight upper and lower bounds on the expected cover time of a random walk on an undirected graph with colored edges. We show that, in general, graphs with two colors have exponential expected cover time, and graphs with three or more colors have doubly-exponential expected cover time. We also give polynomial bounds on the expected cover time in a number of interesting special cases. We describe applications of our results to understanding the dominant eigenvectors of products and weighted averages of stochastic matrices, and to problems on time-inhomogeneous Markov chains.

1 Introduction

A colored graph is a set of n nodes with k distinctly-colored sets of undirected edges. Let the colors be represented by the numbers $1, 2, \ldots, k$. An infinite sequence $C = C_1C_2C_3\ldots$ over alphabet $\{1, 2, \ldots, k\}$ directs a *random walk* on a colored graph from a fixed start node in the following way. At the ith step, a random edge, chosen according to the uniform distribution on edges of color C_i incident to the current node, is followed.

We say that a colored undirected graph G can be *covered from s* if, on every infinite sequence of colors C, a random walk on C starting at s visits every node with probability one. The *expected cover time of G* is defined to be the largest (supremum), over all infinite sequences C and start nodes s, of the expected time to cover G on C starting at s. In this paper we study the expected cover time of colored undirected graphs. Throughout we only consider those graphs that can be covered starting from any node. This property is needed since without it there is no bound on the cover time.

*Supported by NSF Grant CCR91-00886. Computer Science Department, University of Wisconsin, Madison, WI 53706.

†Supported in part by NSF Grant CCR88-13632 and a Lockheed Graduate Fellowship. Computer Science Division, University of California, Berkeley, CA 94720.

In Section 2 we prove the following general bounds on the expected cover time of colored undirected graphs:

- Exponential upper and lower bounds on the expected time to cover undirected graphs with two colors.
- Doubly-exponential upper and lower bounds on the expected time to cover undirected graphs with three or more colors.

In Section 3 we investigate the behavior of walks on randomly chosen color sequences. Here a probability distribution is put on the set of colors $\{1, 2, \ldots, k\}$, assigning color j probability α_j. At each step of the walk the color is chosen independently from this distribution. We prove:

- Exponential upper and lower bounds on the expected time to cover colored undirected graphs on a randomly chosen sequence of colors.

In Section 4 we identify several interesting subclasses of colored graphs and restricted walks that have polynomial expected cover time. In what follows we use the following notation. We use $(C_1C_2 \ldots C_l)^\omega$ to denote the sequence consisting of an infinite number of repetitions of the finite color sequence $C_1C_2 \ldots C_l$. For c in $\{1, 2, \ldots, k\}$ we call the graph induced by those edges that are colored c the *underlying graph* of color c. We use A_c to denote the $n \times n$ probability transition matrix for the underlying graph of color c. We obtain the following bounds:

- Polynomial bounds on the expected cover time of colored graphs when the underlying graphs are connected, aperiodic, and have the same stationary distribution.
- Polynomial bounds on the expected time to cover colored graphs on sequences of the form $(C_1C_2 \ldots C_l)^\omega$ when the matrix product $A_{C_1}A_{C_2} \cdots A_{C_l}$ is irreducible and all entries of its stationary distribution are at least $1/\text{poly}(n)$.
- Polynomial bounds on the expected time to cover colored graphs on randomly chosen sequences when the underlying graphs are connected and all entries of the stationary distribution of $\sum_j \alpha_j A_j$ are at least $1/\text{poly}(n)$.

These results have applications to understanding the eigenvectors of products and weighted averages of stochastic matrices. Our results can be used to show that it is possible for the stationary distribution of a product or weighted average of stochastic matrices to contain exponentially small entries, even when all entries of the stationary distributions of the individual matrices are inversely polynomial.

In Section 5 we investigate the question of whether a colored undirected graph is covered with probability one on all infinite sequences. We use a space-bounded analogue of the probabilistically checkable proof systems of [3] to show that the problem of deciding, given a colored undirected graph G and a start node s, whether G is covered from s with probability one on all infinite sequences is:

- Complete for nondeterministic logspace (NL) when the graph has two colors.
- $PSPACE$-complete for graphs with three or more colors.

Our results also have applications to the theory of time-inhomogeneous Markov chains. Seneta [15] devotes two chapters to the following question: Given an infinite sequence of $n \times n$ stochastic matrices, $M_1, \ldots, M_j, \ldots$, is the sequence weakly-ergodic? That is, do the rows of the matrix $M^{(j)} = \prod_{i=1}^{j} M_i$ tend to equality as $j \to \infty$? Natural complexity-theoretic questions arise from this problem when the sequence $M_1, \ldots, M_j, \ldots$ has a finite description. For example, consider a set $A = \{A_1, \ldots, A_k\}$ of $n \times n$ stochastic matrices. Is it the case that all infinite sequences over the set A are weakly-ergodic? If so, can one bound the rate of convergence to ergodicity for such a sequence as a function of n? The former question was studied in a series of papers [4] [13] [17] motivated by problems from coding theory on finite state channels. In Section 5 we extend our results to show the following:

- It is *PSPACE*-complete to decide weak-ergodicity on sets of two or more stochastic matrices.
- The rate of convergence to ergodicity is doubly-exponential in the worst case.

2 General Bounds on the Cover Time

In this section we present tight bounds on the expected cover time of colored undirected graphs. We show that the expected cover time of a two-colored undirected graph is $2^{\Theta(poly(n))}$, whereas the expected cover time of an undirected graph with three or more colors is $2^{2^{\Theta(n)}}$. We first present the upper bounds in Theorems 2.1 and 2.2 and then present the lower bounds in Theorems 2.3 and 2.4.

We use the following terminology and notation throughout. Let G be a colored undirected graph and let s and t be two nodes of G. We say that t *is reachable from* s *on the color sequence* $C = C_1 \ldots C_l$, if there is a sequence of nodes $s = v_0, v_1, \ldots, v_l = t$ such that G contains an edge of color C_i between v_{i-1} and v_i, for $1 \leq i \leq l$. We call $v_0, v_1, \ldots, v_l$ a *path from* s *to* t *on* C.

Theorem 2.1 *Let G be a colored undirected graph with n nodes. Then the expected cover time of G is $2^{2^{O(n)}}$.*

Proof: Suppose that G can be covered from node s. Fix a color sequence $C_1C_2C_3\ldots$ We consider the random walk on this sequence from node s in intervals of $l = 2^n$ steps. Consider an arbitrary ordering $1, \ldots, n$ of the nodes of G. Suppose that in the first i intervals nodes

$1, \ldots, t-1$ have been visited but t has not been visited. We will show that node t is visited with probability at least $1/n^l$ in the $(i+1)$st interval. Thus, the expected number of intervals after the ith interval until node t is visited is at most n^l. Hence, the expected number of intervals until all nodes are visited is at most nn^l. Since each interval consists of $l = 2^n$ steps, the total expected time needed to cover G from s is at most $n2^n n^{2^n} = 2^{2^{O(n)}}$.

We now show that node t is visited with probability at least $1/n^l$ in interval $(i+1)$, given that it has not been visited in the first i intervals. Let s_i be a node reachable from s in exactly il steps, given that t has not been reached in the first i intervals. It is sufficient to show that node t is reachable from s_i in interval $i+1$. Or, equivalently, that t is reachable from s_i on color sequence $C_{il+1} \ldots C_{il+l'}$, for some $l' \leq l$. If this is the case, the probability that node t is visited, given that s_i is the node reached in il steps, is at least $1/n^l$. This is because at each of the first l' steps of interval $i+1$, with probability at least $1/n$, the path to node t is followed from s_i.

Suppose to the contrary that node t is not reachable from s_i in interval $i+1$. Let $S_0 = \{s_i\}$ and, for $1 \leq m \leq l$, let S_m be the set of nodes reachable from s_i on the color sequence $C_{il+1}C_{il+2} \ldots C_{il+m}$. Since each set S_m is a subset of $\{1, \ldots, n\}$, by the pigeonhole principle $S_j = S_{j'}$ for some $0 \leq j < j' \leq l$. Now consider the color sequence $C_1 C_2 \ldots C_{il+j}(C_{il+j+1} \ldots C_{il+j'})^\omega$. On this sequence with positive probability node t is never reached from s. This is because with positive probability node s_i is reached in exactly il steps on a path that does not visit t, and then node t is not reached in further steps since the reachable nodes are those in S_m, $1 \leq m \leq j'$. This contradicts our assumption that G can be covered from s.

Later we'll show that this bound is tight for graphs with three or more colors. Undirected two-colored graphs, however, are covered in expected time $2^{poly(n)}$. We give a proof of this now.

Theorem 2.2 *Let G be a two-colored undirected graph with n nodes. Then the expected cover time of G is $2^{poly(n)}$.*

Proof: Suppose that G can be covered from node s. Fix a color sequence $C_1 C_2 C_3 \ldots$, where the two colors are red and blue, denoted R and B, respectively. As in Theorem 2.1 we consider the random walk from node s on this sequence in intervals. In this case the intervals are of length $l = (4n-3)(n-1)$. Consider an arbitrary ordering $1, \ldots, n$ of the nodes of G. Suppose that in the first i intervals nodes $1, \ldots, t-1$ have been visited but t has not been visited. We will show that node t is visited with probability at least $1/n^l$ in the $(i+1)$st interval. From this the theorem follows in a manner similar to that of Theorem 2.1.

Again as in Theorem 2.1 it is sufficient to show that node t is reachable from s_i in interval $i+1$, given that t was not visited in the first i intervals, where s_i is a node reachable from s in exactly il steps. This is equivalent to showing that node t is reachable from s_i on the sequence $C_{il+1}C_{il+2} \ldots C_{il+l'}$, for some $l' \leq l$. To keep the notation simple we prove this in the case that $i = 0$, in which case $s_i = s$. The argument is identical for $i \geq 1$.

We first consider the case when the sequence $C_1 C_2 \ldots C_l$ is a prefix of one of the following four strings: $(R)^\omega$, $(B)^\omega$, $(RB)^\omega$, $(BR)^\omega$, and then extend the argument to arbitrary sequences.

Lemma 2.1 *Node t is reachable from s on a prefix of the sequence $(R)^\omega$ (and $(B)^\omega$) of length at most $n-1$.*

Proof: Follows from the fact that the underlying red (blue) graph is connected.

Lemma 2.2 *Node t is reachable from s on a prefix of the sequence $(RB)^\omega$ (and $(BR)^\omega$) of length at most $2n-1$.*

Proof: We prove the lemma for the sequence $(RB)^\omega$. The argument for the sequence $(BR)^\omega$ is analogous. Since G is covered with probability one on all sequences, t is reachable from s on some prefix of $(RB)^\omega$. Consider the shortest path from s to t on a prefix of $(RB)^\omega$. On this path each node of G appears at most once in an even numbered position and once in an odd numbered position. Hence, t is reachable from s on a prefix of $(RB)^\omega$ of length at most $2n-1$.

We now extend the argument to arbitrary sequences $C_1 \ldots C_l$ over $\{R, B\}$. To do this we relate arbitrary color sequences to prefixes of the four strings above using the infinite line graph L shown in Figure 1. Alternate edges of this graph are colored R and B. Thus any sequence of colors defines a unique path from any fixed starting point p on the line. For clarity we will refer to the nodes of L as *points* to distinguish them from the *nodes* of G.

We say that two finite color sequences C and C' are *similar* if starting from any given point on the line L, the unique point reached on the color sequence C is the same as the unique point reached on C'. The following lemma is the key to extending Lemmas 2.1 and 2.2 to arbitrary color sequences.

Lemma 2.3 *Suppose that C is similar to C', where C' is a prefix of $(RB)^\omega$ (or $(BR)^\omega$), and let x and y be nodes of G. If y is reachable from x on C', then y is reachable from x on C.*

Proof: Suppose that from a point p on the line L, point q is reached on the sequences C and C'. Since in the graph G node y is reachable from node x on color sequence C', C' defines a line embedded in G from x to y, along which edges are colored the same as the edges from p to q in L. On color sequence C we construct a path from x to y in the graph G that wanders along this embedded line in the same way that the path from p to q on the sequence C wanders along the line L. Of course the path from p to q on C may visit nodes that do not lie between p and q. In constructing our path from x to y we need to extend our embedded line in G accordingly.

We now make this precise. Let $x = x'_0, x'_1, \ldots, x'_{m'} = y$ be a path from x to y in G and let $p = p'_0, p'_1, \ldots, p'_{m'} = q$ be the path from p to q in L, both on the sequence $C' = C'_1 C'_2 \ldots C'_{m'}$. Let $p = p_0, p_1, \ldots, p_m = q$ be the path from p to q in L on the sequence $C = C_1 \ldots C_m$. We construct a path $x = x_0, x_1, \ldots, x_m = y$ in G on the sequence $C = C_1 \ldots C_m$.

The path is defined inductively as follows. We let $x_0 = x$. Suppose $0 < j \leq m$ and that $x_0, \ldots, x_{j-1}$ are defined. Then x_j is defined as follows:

$$x_j = \begin{cases} x_i, & \text{if } p_j = p_i, \text{ for some } i < j \\ x'_i, & \text{if } p_j = p'_i \\ z, & \text{otherwise, where } z \text{ is any node connected to } x_{j-1} \text{ by an edge of color } C_j. \end{cases}$$

We now continue the proof that t is reachable from s on $C_1 \ldots C_{l'}$, for some $l' \leq l$. Consider the unique path in the line L from any fixed point p on the sequence $C_1 \ldots C_l$. By our choice of $l = (4n-3)(n-1)$ it must be the case that (i) some point of L is visited n times on the sequence $C_1 \ldots C_l$, or (ii) $2n-1$ distinct points to the right of p or to the left of p are visited on the sequence $C_1 \ldots C_l$. In the next two lemmas we show that in either case t is reachable from s on $C_1 \ldots C_{l'}$, for some $l' \leq l$.

Lemma 2.4 *Suppose that some point of L is visited n times on the sequence $C_1 \ldots C_l$. Then t is reachable from s on a prefix of $C_1 \ldots C_l$.*

Proof: Suppose that a point q in L is visited n times on $C = C_1 \ldots C_l$. Then we traverse the red edge adjacent to q (in either direction) at least $n-1$ times, or we traverse the blue edge adjacent to q at least $n-1$ times. Without loss of generality assume that the red edge is traversed $n-1$ times. (The argument in the case that the blue edge is traversed $n-1$ times is analogous.)

Let $s = v_0, v_1, \ldots, v_{m-1}, v_m = t$ be a path from s to t in the underlying red graph, where $m \leq n-1$. We will incorporate this path into a walk on C. Because the red edge adjacent to q is traversed at least $n-1$ times we can rewrite C as follows:

$$C = C^{(0)}RC^{(1)}RC^{(2)}R \cdots C^{(m-1)}RC^{(m)},$$

where $C^{(0)}, C^{(1)}, \ldots, C^{(m-1)}$ are (possibly empty) strings over $\{R, B\}$ that are similar to the empty string ϵ, and $C^{(m)}$ is a string over $\{R, B\}$. Since $C^{(i)}$ $(0 \leq i \leq m-1)$ is similar to ϵ, by Lemma 2.3 for any node x in G there is a path from x back to x on $C^{(i)}$. So on $C^{(i)}$ we can walk from v_i back to v_i, and on the R between $C^{(i)}$ and $C^{(i+1)}$ we can traverse the red edge connecting v_i and v_{i+1}.

Lemma 2.5 *Suppose that $2n-1$ distinct points to the right (or left) of p are visited on the sequence $C_1 \ldots C_l$. Then t is reachable from s on a prefix of $C_1 \ldots C_l$.*

Proof: We do the proof for the case that $2n-1$ distinct points to the right of p are visited and the edge from p to the point to its right is colored R. By Lemma 2.2, on some prefix $C' = C'_1 \ldots C'_m$ of $(RB)^\omega$, where $m \leq 2n-1$, t is reachable from s in G. Let q be the point reachable from p in L on the color sequence $C'_1 \ldots C'_m$. Since $2n-1$ points to the right of p are visited on the sequence $C_1 \ldots C_l$, the point q is reached from p the sequence $C = C_1 \ldots C_{l'}$, for some $l' \leq l$. Thus the sequences C and C' are similar. So by Lemma 2.3 t is reachable from s on $C_1 \ldots C_{l'}$ as required.

In Theorems 2.3 and 2.4 we show that the bounds of Theorems 2.2 and 2.1 are tight. The proofs are based on the following lemma. Before stating the lemma we need the following generalization of a strongly-connected directed graph.

A k-colored directed graph G is *strongly-connected* if for every infinite sequence of colors C over $\{1, 2, \ldots, k\}$, and every pair of nodes u and v, v is reachable from u on a prefix of C. Note that a strongly-connected colored graph is covered with probability one on all infinite sequences from all starting nodes.

Lemma 2.6 *For every strongly-connected k-colored directed graph G there is a $(k+1)$-colored undirected graph G' such that:*

1. *the number of nodes in G' is twice the number of nodes in G,*
2. *G' can be covered from all its nodes, and*
3. *for every k-color sequence C, there exists a $(k+1)$-color sequence C' such that the expected cover time of G' on C' is twice the expected cover time of G on C.*

Proof: Suppose that G is a strongly-connected k-colored directed graph with nodes $\{u_1, \ldots, u_n\}$ and edges colored $\{1, 2, \ldots, k\}$. Let G' have nodes $\{v_1, \ldots, v_n, w_1, \ldots, w_n\}$, with an edge colored $k+1$ connecting v_i and w_i, for all i. For each directed edge from u_i to u_j in G, make an undirected edge between w_i and v_j in G' of the same color. Also make a complete graph on $\{w_1, w_2, \ldots, w_n\}$ in color $k+1$, and complete graphs on $\{v_1, v_2, \ldots, v_n\}$ in each of colors $1, 2, \ldots, k$.

The lemma is a direct consequence of the following two facts, both of which are routine to verify.

1. G' can be covered from all its nodes.
2. For every walk on G starting from u_i on color sequence $C_1 C_2 \ldots C_l$, there is a corresponding walk on G' starting from v_i on color sequence $(k+1)C_1(k+1)C_2 \ldots (k+1)C_l(k+1)$. The two walks have the same probability, and node u_j is visited in the first walk if and only if nodes v_j and w_j are visited in the second.

Lemma 2.6 shows how to simulate a random walk in a k-colored directed graph with a random walk in a $(k+1)$-colored undirected graph. We use the construction to prove the following two lower bounds on the expected cover time of colored undirected graphs.

Theorem 2.3 *There are two-colored undirected graphs that are covered from all nodes and have expected cover time* $2^{\Omega(n)}$.

Proof: We obtain this bound by applying the construction of Lemma 2.6 to a family of strongly-connected directed graphs with exponential expected cover time. An example of such a family of graphs is given by a sequence of nodes numbered $1, 2, \ldots, n$ with a directed edge from node i to node $i+1$, for $1 \leq i \leq n-1$, and a directed edge from node i to node 1, for $2 \leq i \leq n$.

Theorem 2.4 *There are three-colored undirected graphs that can be covered from all nodes and have expected cover time* $2^{2^{\Omega(n)}}$.

Proof: In [7] Condon and Lipton construct a family of strongly-connected two-colored directed graphs with $O(n)$ nodes and expected cover time $2^{2^{\Omega(n)}}$. On a particular sequence of colors a random walk on the nth graph in the family simulates 2^n tosses of a fair coin and reaches an absorbing state if and only if all outcomes were heads. The bound is obtained by applying the construction of Lemma 2.6 to that family of graphs.

3 Graphs with Self-Loops and Random Color Sequences

In this section we strengthen the exponential lower bound of Theorem 2.3 on the expected cover time of undirected two-colored graphs. We consider *graphs with self-loops* in which there is a self-loop of each color at each node. Note that if all of the underlying graphs in a graph with self-loops are connected, then the graph can be covered from any node. This is because for all nodes s and t, and all color sequences C of length $2n$, t is reachable from s on C.

It might seem that graphs with self-loops have polynomial expected cover time. Certainly if a self-loop of each color is added to each node of the graph of Theorem 2.3, the resulting graph has polynomial expected cover time. In the following theorem we show that this does not happen in general. We prove that the expected cover time of graphs with self-loops is exponential, strengthening the result of Theorem 2.3.

The theorem strengthens Theorem 2.3 in another way. It shows that the expected cover time is exponential, even on a random sequence. This fact, together with the results of the next section, has applications in understanding the eigenvectors of weighted averages of stochastic matrices.

Theorem 3.1 *There are colored undirected graphs with self-loops that have expected cover time* $2^{\Theta(n)}$ *on the sequence* $(RB)^{\omega}$*, and on a randomly chosen sequence of colors.*

Proof: We present in Figure 2 an example of a two-colored graph with self-loops which has exponential expected cover time on a randomly chosen sequence of colors. The solid lines are the red edges and the dotted lines are the blue edges; there is also a self-loop of each color at each node, but they have been left out of the diagram.

We show that the expected time to reach node n of the graph from node 1 is exponential in n on a random sequence of colors. In what follows we call nodes $1, \ldots, n$ the *primary* nodes, and nodes $1', \ldots n'$ the *secondary* nodes. Suppose a random walk from i is performed on a random sequence of colors until a primary node other than i is reached. This primary node must be either $i+1$ or $i-1$. Let $p(i, i+1)$ be the probability that the next primary node reached is $i+1$. $p(i, i-1)$ is defined similarly.

The construction ensures that for $2 \leq i \leq n-1$, $p(i, i+1) = 1/2 - \epsilon$, where ϵ is a positive constant that is independent of i.

To get an intuitive understanding of why $p(i, i+1) < p(i, i-1)$, observe that the walk from primary node i to primary nodes $i+1$ and $i-1$ may or may not go through a secondary node. The last edge on a direct path (one that is not completed via a secondary node) to primary node $i+1$ goes through an edge that is one of four of the same color, whereas the corresponding path to primary node $i-1$ goes through an edge that is one of three of the same color. The color at each step, however, is decided by the toss of a fair coin. On the other hand, if secondary node $(i-1)'$ is reached, primary node $i-1$ is much more likely than primary node i to be next. But if secondary node i' is reached, primary node i is much more likely than primary node $i+1$. In fact, a brute force calculation of the probabilities shows that $p(i, i+1) = 35/78$.

We use this to prove a lower bound on the expected time to reach node n from node 1 on a random sequence of colors. We define a *superstep* as a walk that begins at some primary node i and ends as soon as primary node $i+1$ or $i-1$ is reached. Since each superstep takes at least one step of the random walk, a lower bound on the expected number of supersteps is a lower bound on the expected number of steps of the random walk.

Let $T(i, i+1)$ be the expected number of supersteps to reach node $i+1$ from node i. Then $T(n-1, n)$ is a lower bound on the expected time to reach n from 1. Clearly $T(i, i+1)$ satisfies the following recurrence:

$$\begin{aligned} T(i, i+1) &= p(i, i+1) + (1 - p(i, i+1))(1 + T(i-1, i) + T(i, i+1)) \\ T(1, 2) &= 1 \end{aligned}$$

The solution to this recurrence shows that $T(i, i+1) \geq ((1-p)/p)^{i-1}$, where $p = p(i, i+1)$. Hence, $T(n-1, n) \geq c^{n-2}$, where $c = (1/2 + \epsilon)/(1/2 - \epsilon) > 1$.

This argument shows the existence of a sequence on which the expected cover time is exponential. The proof that $(RB)^{\omega}$ is one such sequence is straightforward but tedious; we omit it here.

4 Polynomial Special Cases

In this section we identify several conditions under which colored undirected graphs are covered in polynomial expected time. Our results are summarized below.

- (Theorem 4.1) We show that colored graphs are covered in polynomial expected time if the underlying graphs are aperiodic and have a common stationary distribution.
- (Theorems 4.2 and 4.4) We also consider the expected cover time of colored graphs on sequences of the form $(C_1C_2\ldots C_l)^\omega$, where l is a constant. We show that the expected cover time of colored graphs on such sequences is polynomial if the product $A_{C_1}A_{C_2}\cdots A_{C_l}$ is irreducible and all entries of its stationary distribution are at least $1/\text{poly}(n)$.
- (Theorem 4.3) We also consider the expected cover time of colored graphs on randomly chosen sequences, where at each step of the walk color j is chosen with probability α_j. We show that the expected cover time of colored graphs on such sequences is polynomial if all entries of the stationary distribution of $\sum_j \alpha_j A_j$ are at least $1/\text{poly}(n)$.

We use the following notation in this section. Let c be a color. We use E_c to denote the set of edges of color c. For a node i, let $N_c(i)$ denote the set of neighbors of i along edges of color c and let $d_c(i)$ be $|N_c(i)|$. Let A_c denote the $n \times n$ stochastic matrix whose $\{i,j\}$th entry is the probability of reaching j from i in one step, when an edge of color c is followed. Then the $\{i,j\}$th entry of A_c is $1/d_c(i)$ if there is an edge of color c connecting i and j, and 0 otherwise. Let π_c be an n-vector satisfying $\pi_c = \pi_c A_c$. If the underlying graph colored c is connected, π_c is the unique vector of stationary probabilities and has ith entry $d_c(i)/2|E_c|$.

In the following theorem we show that colored graphs are covered in polynomial expected time if the underlying graphs are aperiodic and have the same stationary distribution.

Theorem 4.1 *Let G be a colored undirected graph with n nodes which is is connected in each color. If the underlying graphs are aperiodic and have the same stationary distribution, then the expected cover time of G is $O(n^5 \log n)$.*

Proof: Let π be the common stationary distribution of the underlying graphs. Suppose for now that our color sequence is $(C_1)^\omega$; that is, that we are taking a random walk on an aperiodic undirected graph. We will generalize this later to arbitrary sequences.

Let v_t be the n-vector whose ith entry (denoted $v_t(i)$) is the probability of being at node i after t steps of a random walk starting at j. Let v_0 be the n-vector with a 1 in the jth position and 0's everywhere else. Then $v_t = (A_{C_1})^t v_0$ and, as $t \to \infty$, $v_t \to \pi$. Let Δ_t be the discrepancy vector at time t, defined as $\Delta_t = v_t - \pi$, and let $\| \Delta_t \| = \sum_i \Delta_t^2(i)$. Then $\| \Delta_t \|$ measures the distance of v_t from π, so a bound on the rate at which $\| \Delta_t \|$ approaches 0 gives a bound on the rate at which v_t approaches π.

Results of Alon [2], Jerrum-Sinclair [11], and Mihail [12] show that for t polynomial in n $\| \Delta_t \| \leq 1/\exp(n)$. The exact polynomial depends on the cutset expansion of the graph and is at most $O(n^3)$. The proof in [12] shows this by obtaining an appropriate lower bound on $\| \Delta_t \| - \| \Delta_{t+1} \|$, the amount by which the discrepancy drops in one time step. This bound depends only on $\| \Delta_t \|$ and the probability matrix A_{C_1} and, in particular, does not depend on how the discrepancy $\| \Delta_t \|$ was arrived at. The incremental nature of this argument makes it readily applicable to random walks on arbitrary sequences.

If $C_1C_2C_3\ldots$ is the color sequence, let v'_t be the probability vector for a random walk on $C_1C_2\ldots C_t$ starting at j and let Δ'_t be the discrepancy at time t. Then $v'_t = A_{C_t}\cdots A_{C_2}A_{C_1}v_0$ and $\Delta'_t = v'_t - \pi$. Applying the previous results we get that for $t = O(n^3)$, $\| \Delta'_t \| \leq 1/\exp(n)$. By definition, $\pi(i) \geq 1/n^2$ for all i, so $v'_t(i) \geq 1/cn^2$, where c is a positive constant.

¿From this we derive bounds on the expected cover time by viewing the process as a coupon collector's problem on cn^2 coupons, where sampling one coupon takes $O(n^3)$ steps of a random walk. This analysis gives an $O(n^5 \log n)$ bound on the expected cover time.

An extension of this argument shows that the aperiodicity requirement can be somewhat relaxed, while still obtaining the same bound. If the underlying graphs have the same stationary distribution and some, or all, of them are bipartite, the graph can still be covered in polynomial expected time, provided that the bipartitions in the underlying bipartite graphs are the same. It is an open question whether the expected cover time is polynomial when the bipartitions do not all coincide.

In the following theorem we show that colored undirected graphs are covered in polynomial expected time on sequences of the form $(C_1C_2\ldots C_l)^\omega$, if the product $A_{C_1}A_{C_2}\cdots A_{C_l}$ is irreducible and all entries of its stationary distribution are at least $1/\text{poly}(n)$.

Theorem 4.2 *Let G be a colored undirected graph with n nodes which is connected in each color, and let $C_1C_2\ldots C_l$ be a sequence of colors, for some constant l. Suppose that the matrix product $A_{C_1}A_{C_2}\cdots A_{C_l}$ is irreducible, and that all entries of its stationary distribution π are at least $1/p(n)$, for some polynomial $p(n)$.*

Then the expected cover time of G on the sequence $(C_1C_2\ldots C_l)^\omega$ is $O(n^{l+2}p(n))$.

Proof: Let G_P be the weighted directed graph with n nodes and probability transition matrix $P = A_{C_1}A_{C_2}\cdots A_{C_l}$. In what follows we show that the expected cover time of G_P is at most $2n^{l+2}p(n)$. This implies that the expected cover time of G on $(C_1C_2\ldots C_l)^\omega$ is at most $2ln^{l+2}p(n)$.

Since P is irreducible, there is a directed walk on G_P from any starting node that visits every node at least once and has length at most n^2. We bound the expected time for a random walk on G_P to complete such a walk.

Let i and j be a pair of nodes in G_P such that $P_{ij} > 0$. We bound the expected time for a random walk that begins at i to traverse the edge from i to j.

Each time the walk is at node i it traverses the edge from i to j with probability P_{ij}. Hence, the expected number of returns to i until the edge from i to j is traversed is $1/P_{ij}$. If $P_{ij} = 1$, the expected time to traverse the edge from i to j is 1, and we are done. In what follows we assume that $0 < P_{ij} < 1$.

Let $T(i,i)$ denote the mean recurrence time of node i. Then the expected time to return to i, given that the edge from i to j is not traversed, is at most $T(i,i)/(1 - P_{ij})$. Hence, the expected time for the walk to traverse the edge from i to j is at most $T(i,i)/P_{ij}(1 - P_{ij})$.

Since $P = A_{C_1}A_{C_2}\cdots A_{C_l}$ and each non-zero entry of the A_{C_i} is at least $1/n$, each non-zero entry of P is at least $1/n^l$. Also $1 - P_{ij}$ is at least $1/n^l$. Hence, $P_{ij}(1-P_{ij}) \geq (1/n^l)(1-1/n^l) \geq 1/2n^l$, and the expected time for the walk to traverse the edge from i to j is at most $2n^l T(i,i)$. Then, from the fact that the mean recurrence time of node i is the reciprocal of its stationary probability $\pi(i)$, we get that the expected time for the walk to traverse the edge from i to j is at most $2n^l p(n)$. It follows that the expected time to cover G_P is at most $2n^{l+2}p(n)$.

Together Theorems 4.2 and 3.1 have the following interesting interpretation. They show that, in general, the stationary distribution of a product of stochastic matrices can contain exponentially small entries, even when the entries of the stationary distributions of the individual matrices are bounded below by $1/\text{poly}(n)$.

Recall the graph G constructed in Theorem 3.1. G has self-loops, so the product matrix $A_R A_B$ is irreducible. Since A_R and A_B correspond to undirected graphs, all entries in their stationary distributions are inversely polynomial. But the expected cover time of G on $(RB)^\omega$ is exponential. It follows from Theorem 4.2 that at least one entry in the stationary distribution of $A_R A_B$ is exponentially small.

Let G_P be any weighted, directed graph and let P be its probability transition matrix. The key idea in Theorem 4.2 is that if (1) P is irreducible, (2) all non-zero entries of P are at least $1/\text{poly}(n)$, and (3) all entries of the unique stationary distribution of P are at least $1/\text{poly}(n)$, then the expected cover time of G_P is polynomial in n. We apply this idea again to obtain a similar result about the expected cover time on a randomly chosen color sequence.

Theorem 4.3 *Let G be a colored undirected graph with n nodes which is connected in each of its k colors, and let $\alpha_1, \alpha_2, \ldots, \alpha_k$ be constants such that $0 \leq \alpha_i \leq 1$ and $\sum_i \alpha_i = 1$. Suppose that all entries of the stationary distribution of the matrix $\sum_i \alpha_i A_i$ are at least $1/p(n)$, for some polynomial $p(n)$.*

Then the expected cover time of G on a randomly chosen sequence of colors, where at each step color i is chosen independently with probability α_i, is $O(n^2 p(n))$.

Proof: The matrix $P = \sum_i \alpha_i A_i$ is irreducible and has all entries at least $1/kn$. Having made this observation the rest of the proof is analogous to that of Theorem 4.2.

Together Theorem 4.3 and the graph of Theorem 3.1 show that the stationary distribution of a weighted average of stochastic matrices can contain exponentially small entries, even when the entries of the stationary distributions of the individual matrices are bounded below by $1/\text{poly}(n)$.

We conclude this section by remarking that a more refined argument based on the techniques of Aleliunas et al. [1] and Göbel and Jagers [9] improves upon Theorems 4.1 and 4.2 in the following special case. We omit the proof here.

Theorem 4.4 *Let G be a colored undirected graph with n nodes which is connected in each color, and let C_1, C_2 be a pair of colors. Suppose that matrices A_{C_1} and A_{C_2} have the same stationary distribution and that the product $A_{C_1}A_{C_2}$ is irreducible.*

Then the expected cover time of G on the sequence $(C_1C_2)^\omega$ is $O(n\min\{|E_{C_1}|, |E_{C_2}|\})$.

5 Complexity Results and Applications

In our proofs bounding the expected cover time of colored undirected graphs we assume that the graphs are covered with probability one on all infinite sequences. A natural question asks about the complexity of deciding whether a given colored undirected graph satisfies this condition. We begin this section by showing that the problem of deciding, given a colored undirected graph G and a start node s, whether G is covered from s with probability one on all infinite sequences is: (1) complete for nondeterministic logspace (NL) when the graph has two colors, and (2) *PSPACE*-complete for graphs with three or more colors.

Theorem 5.1 *The problem of deciding, given a two-colored undirected graph G and a start node s, whether G is covered from s with probability one on all infinite sequences is NL-complete.*

Proof: We first describe a nondeterministic logspace Turing machine for deciding, given a two-colored undirected graph G and a start node s, whether G is *not* covered with probability one on all infinite sequences. It follows that the stated problem is in NL since $NL = coNL$ [10] [16].

We use the following equivalence, the proof of which is implicit in the proof of the exponential upper bound of Theorem 2.2.

A walk from s visits node t with probability strictly less than one on some infinite color sequence if and only if there exists a node v such that:

(1) v is reachable from s on a path of finite length that avoids t, and

(2) t is not reachable from v on at least one of $(R)^\omega$, $(B)^\omega$, $(RB)^\omega$, $(BR)^\omega$.

Notice that the color sequence for the path from s to v is not specified, so the path can be restricted to have length at most $n-2$. Also recall that paths on $(R)^\omega$ and $(B)^\omega$ have length at most $n-1$, and paths on $(RB)^\omega$ and $(BR)^\omega$ have length at most $2n-1$. A nondeterministic logspace Turing machine can guess t, v, and a path of length at most $n-2$ from s to v that avoids t. It can also guess which of the four types of string from (2) above does not admit a path from v to t. Since such a path, if it exists, has length $O(n)$, the techniques of Immerman [10] can be used to verify that there is no path from v to t on the guessed string.

For the hardness result consider the following decision problem. Given a directed graph G and a node s, is it the case that an infinite random walk beginning at s covers G with probability one? This problem is NL-complete by reduction from s-t connectivity. (Take the instance of s-t connectivity and add edges (t, v) and (v, s), for all v.) The rest of the proof follows by applying the construction of Lemma 2.6 to this problem.

Next we show that the analogous problem for graphs with three or more colors is $PSPACE$-complete. For the hardness result, we extend previous results on space-bounded interactive proof systems. We define a space-bounded analogue of the probabilistically checkable proof systems of [3].

A *verifier* is a probabilistic Turing machine that takes as input a pair (x, π), where x and π are strings over the alphabet $\{0, 1\}$. The string π is called a *proof*, and can be infinitely long. The proof π is stored on a one-way infinite, read-only tape. The verifier is constrained to read π in one direction; in fact, the head on π begins at its leftmost symbol and moves right in every step. The string x is also stored on a read-only tape, but its length is finite, and the verifier can read x in both directions. Let n denote the length of x. A language L is in $PCP(\log n)$ if there exists an $O(\log n)$ space-bounded verifier V satisfying the following properties:

1. For all $x \in L$, there exists a (finite) proof $\pi \in \{0, 1\}^*$ such that V accepts (x, π) with probability 1.
2. For all $x \notin L$, on all proofs π, V rejects (x, π) with probability $\geq 2/3$.
3. V halts (accepts or rejects) with probability 1 on all inputs (x, π). In fact, starting from any possible setting of its worktape, state, and tape heads, V halts with probability 1.

Adapting previous proofs of Condon [6] and Dwork and Stockmeyer [8], we show that *PSPACE* is contained in $PCP(\log n)$. We then reduce the problem of deciding if an input x is accepted by the verifier of such a proof system to the covering problem for three-colored undirected graphs. For completeness, we include a sketch of the proof that $PSPACE \subseteq PCP(\log n)$ in Appendix A.

Theorem 5.2 *The problem of deciding, given a three-colored undirected graph G and a start node s, whether G is covered from s with probability one on all infinite sequences is PSPACE-complete.*

Proof: We show that the problem is in $PSPACE$ by describing a nondeterministic polynomial space-bounded Turing machine for deciding, given a colored undirected graph G and a start node s, whether G is *not* covered with probability one on all infinite sequences. It follows that the original problem is in $PSPACE$ since $PSPACE$ is closed under complement [10] [16] and under the addition of nondeterminism [14].

We use the following equivalence, the proof of which is implicit in the proof of the doubly-exponential upper bound of Theorem 2.1.

A walk from s visits node t with probability strictly less than one on some infinite color sequence if and only if there exists a node v such that:

(1) v is reachable from s on a path of finite length that avoids t, and

(2) for some color sequence C of length 2^n, t is not reachable from v on any prefix of C.

Again the color sequence for the path from s to v is not specified, so the path can be restricted to have length at most $n-2$. A nondeterministic polynomial space-bounded Turing machine can guess t, v, and a path of length at most $n-2$ from s to v that avoids t. It can also guess C one character at a time and verify that t is not reachable from v on each successive prefix of C.

For the hardness result, we show that the computation of a logspace verifier V on input x can be represented by a two-colored *directed* graph G_x as follows. The nodes of G_x correspond to the configurations of V on input x. A configuration encodes V's state, the contents of its worktape, the head position on the worktape, and the head position on x (but not the head position on π). We assume that the nodes are numbered, and that *start*, *reject*, and *accept* denote the numbers of nodes corresponding to the unique starting, rejecting, and accepting configurations, respectively. Since V is $O(\log n)$ space-bounded the number of nodes is poly(n). From each node the blue edges describe the transitions of the verifier if the current symbol of π is 1, and the red edges describe the transitions if the current symbol of π is 0.

We now show that the membership problem for any language L in PCP$(\log n)$ can be reduced to the problem of determining if a two-colored directed graph is covered with probability one on all infinite sequences. Suppose that L is accepted by a verifier V with properties (1)-(3) above. Let G_x be the graph describing the computation of V on x. We also add the following edges to G_x. There is an edge $(accept, z)$ of color c if there is an edge $(start, z)$ of color c. There is also a red edge and a blue edge from *reject* to every other node in G_x, including *start* and *accept*.

If x is not in L then, on every proof π, the rejecting configuration is reached with positive probability. Thus, a random walk on G_x on any sequence of colors eventually reaches *reject* with probability one. From *reject* every other node in G_x is reachable in one step. It follows that G_x is covered with probability one.

On the other hand, if x is in L, then there is a finite proof π that causes V to accept with probability one. On the sequence of colors corresponding to repeating π ad infinitum, *reject* is never reached from *start*. This is because *accept* is repeatedly reached from *start* with probability one.

The rest of the proof comes from converting the two-colored directed graph G_x to a three-colored undirected graph using the construction of Lemma 2.6.

Our results also have applications to the theory of time-inhomogeneous Markov chains. An infinite sequence of $n \times n$ stochastic matrices $M_1, \ldots, M_j, \ldots$, is *weakly-ergodic* if the rows of the matrix $M^{(j)} = \prod_{i=1}^{j} M_i$ tend to equality as $j \to \infty$. Intuitively, a sequence of matrices is weakly-ergodic if the limiting distributions are independent of the starting state. Natural complexity-theoretic questions arise when the matrices of the sequence come from a finite set $A = \{A_1, \ldots, A_k\}$. We show that to decide whether all infinite sequences over a set are weakly-ergodic is $PSPACE$-complete if the set contains at least two matrices. We also show that the rate of convergence to ergodicity is doubly-exponential in the worst case.

Theorem 5.3 *Given a set $A = \{A_1, A_2, \ldots, A_k\}$ of two or more $n \times n$ stochastic matrices it is PSPACE-complete to decide whether all infinite sequences over A are weakly-ergodic.*

Proof: A stochastic matrix M is *ergodic* if the limit of M^t, as $t \to \infty$, exists and has all rows equal. In [13] Paz showed that all infinite sequences over A are weakly-ergodic if and only if all matrices $M = A_{i_1} A_{i_2} \cdots A_{i_l}$, where $1 \leq i_j \leq k$ and $l \leq (3^n - 2^{n+1} - 1)/2$, are ergodic. A nondeterministic polynomial space-bounded Turing machine can guess the indices $i_1, i_2, \ldots, i_l$ of a non-ergodic matrix $M = A_{i_1} A_{i_2} \cdots A_{i_l}$. The machine cannot compute the matrix M since its entries can be doubly-exponentially small, but it can compute and store B, the $n \times n$ matrix whose $\{i, j\}$th entry is 1 if $M_{ij} > 0$, and 0 otherwise. Using the fact that M is ergodic if and only it is irreducible and aperiodic, we can use the matrix B to decide the non-ergodicity of M in nondeterministic polynomial time. Using B we can determine whether M is irreducible in deterministic polynomial time. To decide periodicity of M, we observe that a nondeterministic Turing machine can guess a partition of $\{1, 2, \ldots, n\}$ into $S_0, S_1, \ldots, S_m$, with $m > 0$, and verify that, for all $1 \leq i, j \leq n$, if $i \in S_r$ and $B_{ij} = 1$, then $j \in S_{(r+1) \bmod m}$. This verification procedure can be performed in polynomial time. Since *PSPACE* is closed under complement [10] [16] and the addition of nondeterminism [14] this shows that deciding weak-ergodicity is in *PSPACE*.

For hardness, we show that the membership problem for any language in $PCP(\log n)$ can be reduced to deciding weak-ergodicity. Let G_x be the graph corresponding to the computation of V on x. We add an edge $(accept, z)$ of color c to G_x if there is an edge $(start, z)$ of color c. We also add a red self-loop and a blue self-loop at *reject*, making *reject* a sink. Let A be the set containing the probability transition matrices of the red and blue graphs.

If x is not in L, then a walk on G_x reaches *reject* with probability one on any infinite sequence of colors. Correspondingly, the limit of any infinite product of matrices exists and has all rows equal. Each row of the limiting matrix has a zero in every entry, except for the *reject* entry where there is a one.

On the other hand, if x is in L, then there is a finite proof π that causes V to accept with probability one. Consider a walk on G_x on the sequence of colors corresponding to repeating π ad infinitum. If the walk begins at *reject* it remains at *reject* forever, but if it begins at *start* it never reaches *reject*. So the *reject* column of the *reject* row contains a one, but the *reject* column of the *start* row is zero. Hence, the rows do not tend to equality in this case.

Finally, we remark that the lower bound of Theorem 2.4 implies a doubly-exponential lower bound on the rate of convergence to ergodicity.

6 Concluding Remarks

We give bounds on the expected cover time of colored undirected graphs. We show that, in general, the expected cover time is exponential for two colors, and doubly-exponential for three or more colors. We remark that there is a gap in the bounds for graphs with two colors. The upper bound is based on the fact that the maximum distance between any pair of nodes on any color sequence is $O(n^2)$. In the graph we construct for our lower bound, however, the maximum distance is $\Theta(n)$. This results in an upper bound of $2^{O(n^2)}$ and a lower bound of $2^{\Omega(n)}$. It is an open problem to close this gap.

We identify two properties of the underlying graphs and consider their effect on the expected cover time. The first property is that the underlying graphs are aperiodic, and the second that they all have the same stationary distribution. We show that if both properties are satisfied, the expected cover time is polynomial, and that if neither holds, it is exponential. We show that if the stationary distributions differ even slightly (as in the example of Theorem 3.1) the expected cover time is again exponential, even when the underlying graphs are aperiodic. An open question is whether the expected cover time is polynomial when the stationary distributions are the same, but some of the underlying graphs are periodic. If all of the bipartitions in the underlying periodic graphs are the same, the expected cover time is polynomial, but when the bipartitions do not all coincide the question remains open.

Acknowledgments

Richard Lipton first encouraged us to consider random walks in colored undirected graphs. Thanks to Milena Mihail for several helpful discussions, and to Prasoon Tiwari for his contributions to Theorem 4.4.

References

[1] R. Aleliunas, R. Karp, R. Lipton, L. Lovász, and C. Rackoff. Random walks, universal traversal sequences, and the complexity of maze problems. In *Proc. of 20th Symposium on*

Foundations of Computer Science, 1979.

[2] N. Alon. Eigenvalues and expanders. In *Combinatorica* 6(2):83-96, 1986.

[3] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and hardness of approximation problems. In *Proc. of 33rd Symposium on Foundations of Computer Science*, 1992.

[4] D. Blackwell, L. Breiman and A. J. Thomasian. Proof of Shannon's transmission theorem for finite-state indecomposable channels. In *Annals of Math. Stat.* 29:1209-1220, 1958.

[5] A. Broder and A. Karlin. Bounds on the cover time. In *Journal of Theoretical Probability* 2(1), 1989.

[6] A. Condon. Space-bounded probabilistic game automata. In *JACM* 38(2):472-494, April 1991.

[7] A. Condon and R. Lipton. On the complexity of space-bounded interactive proofs. In *Proc. of 30th Symposium on Foundations of Computer Science*, 1989.

[8] C. Dwork and L. Stockmeyer. Finite state verifiers I: The power of interaction. In *JACM* 39(4):800-828, October 1992.

[9] F. Göbel and A. A. Jagers. Random walks on graphs. In *Stochastic Processes and their Applications*, 2:311-336, 1974.

[10] N. Immerman. Nondeterministic space is closed under complement. In *Proceedings of the Conference on Structure in Complexity Theory*, 1988.

[11] M. Jerrum and A. Sinclair. Conductance and the rapid mixing property for Markov chains. In *Proc. of 20th ACM Symposium on Theory of Computing*, 1988.

[12] M. Mihail. Conductance and convergence of expanders. In *Proc. of 30th Symposium on Foundations of Computer Science*, 1989.

[13] A. Paz. Definite and quasi-definite sets of stochastic matrices. In *AMS Proceedings* 16(4):634-641, 1965.

[14] W. J. Savitch. Relationships between nondeterministic and deterministic tape complexities. In *Journal of Computer and Systems Sciences* 4(2):177-192, 1970.

[15] E. Seneta. *Non-negative Matrices and Markov Chains* (2nd Edition). Springer-Verlag, New York, 1981.

[16] R. Szelepcsényi. The method of forcing for nondeterministic automata. In *Bulletin of the European Association for Theoretical Computer Science*, 1987.

[17] J. Wolfowitz. Products of indecomposable, aperiodic, stochastic matrices. In *AMS Proceedings* 14:733-737, 1963.

A $PSPACE \subseteq PCP(\log n)$

Proof: (sketch) Let L be any language in $PSPACE$, and let M be a Turing machine that accepts L using $p(n)$ space on input x of length n, where p is a polynomial. A *configuration* of M is an encoding of its tape contents, head position, and state at one step of its computation. Without loss of generality, assume that M counts its steps and rejects if it detects looping. Let $k > 0$ be the smallest integer such that $2^k \geq p(n)$. We will pad the tape in a configuration of M to have length exactly 2^k. Note that this at most doubles the length of the tape. We now represent the computation of M on x as a sequence of configurations of length 2^k. We begin the encoding with $k+1$ ones, and each pair of consecutive configurations is separated by $k+1$ ones. The number of configurations in the encoding is bounded by 2^{2^k}, an exponential function in n.

An $O(\log n)$ space-bounded verifier V can check that a given position in a configuration is consistent with the next configuration. The verifier must simply remember $O(1)$ symbols of the configuration and then count to $2^k + k + 1$, advancing through the encoding as it counts. When V has finished counting, it can check the corresponding positions in the next configuration.

The verifier can choose a random position in the configuration to check by tossing k coins while it reads k of the $k+1$ ones that precede the configuration. Notice that if V checks the consistency of configuration $j-1$ with configuration j, then V is unable to check the consistency of configuration j with configuration $j+1$. For this reason V tosses a $(k+1)$st coin, and the outcome tells V whether or not to check the configuration that follows.

The verifier can check that the first configuration is correct; that is, that the computation begins in the start state with x on its tape. The verifier can also check that the last configuration is an accepting configuration. If either of these tests fail, or if the rejecting configuration ever appears, V rejects. If the computation contains an inconsistency in any of the intermediate steps, V detects it with probability at least $1/2^{k+1}$ and rejects.

To reduce the probability of error, we concatenate 2^{k+2} encodings of the computation of M on x. The verifier can count the encodings as it does the consistency checks. If V checks 2^{k+2} computations and no consistency check fails, then V accepts. If π is finite in length and V reaches the end of π without accepting, then V rejects.

If x is in L, then on the proof π which is the encoding of an accepting computation of M on x repeated 2^{k+2} times, V accepts with probability 1.

Suppose that x is not in L, and let π be any proof. If the first $2^k + k + 1$ symbols of π do not encode the starting configuration of M on x preceded by $k+1$ ones, then V rejects. Assume that the starting configuration is correctly encoded, and suppose that the accepting configuration appears 2^{k+2} times in π. Consider π parsed into $\pi_1 \pi_2 \ldots \pi_{2^{k+2}} \pi'$. The string π_1 is the initial portion of π, up to the first occurrence of the accepting configuration. For $2 \leq i \leq 2^{k+2}$, π_i is the portion of π that follows π_{i-1}, up to the ith occurrence of the accepting configuration. The string π' is everything that follows the 2^{k+2}nd occurrence of the accepting configuration in π. Since x is not in L, for all $1 \leq i \leq 2^{k+2}$, there is an inconsistency in the computation encoded by π_i. So, for all $1 \leq i \leq 2^{k+2}$, V detects an inconsistency in π_i and rejects with probability at least $1/2^{k+1}$. Hence, the probability that V accepts is at most $(1 - 2^{-(k+1)})^{k+2} \leq 1/3$.

Suppose that the accepting configuration appears fewer than 2^{k+2} times in π. Let π' be all of π after the last occurrence of the accepting configuration. If π' is finite or if π' contains the rejecting configuration, then V rejects. Suppose that π' is infinite and does not contain the rejecting configuration. Consider π' in pieces of length $(2^{2^k} + 1)(2^k + k + 1)$. Since M counts its steps and rejects if it loops, each such piece contains an inconsistency. In each piece the verifier detects an inconsistency and rejects with probability at least $1/2^{k+1}$. Hence, V rejects with probability one in this case.

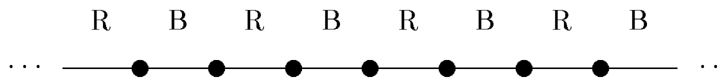

Figure 1: Line graph L

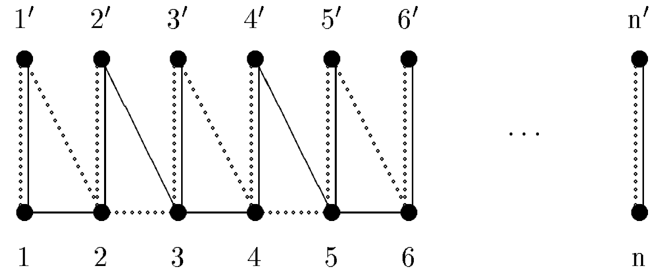

Figure 2: Exponential time graph with self-loops