Random Walks in Colored Graphs

Anne Condon [*] Diane Hernek [†]

Abstract

We give tight upper and lower bounds on the expected cover time of a random walk in an undirected graph with colored edges. We show that for graphs with two colors the expected cover time is exponential, and that for three or more colors it is double exponential. In addition, we give polynomial bounds in a number of interesting special cases. We describe applications of these results to understanding the eigenvalues of products and weighted averages of matrices, and to problems on time-inhomogeneous Markov chains.

1 Introduction

We introduce the notion of a random walk in an undirected colored graph and analyze the expected cover time of such walks. A colored graph is a set of n nodes with k distinctly-colored sets of undirected edges. The colors are $\{1, 2, \ldots k\}$. An infinite sequence $C = C_1C_2C_3\ldots$, where each C_i is in $\{1, 2, \ldots k\}$, defines a *random walk* on a colored graph from a fixed start node s in the following way. At the i-th step, an edge, chosen randomly and uniformly from the edges of color C_i at the current node, is followed.

We say that G can be *covered from* s if, on every infinite sequence C of colors, a random walk on C starting at s visits every node with probability one. The *expected cover time of* G *on* C is defined as the maximum, over all nodes s of G, of the expected time to cover G from s on C. The *expected cover time* of G is defined to be the supremum, over all infinite sequences of colors C, of the expected cover time of G on C.

In this paper we study the expected cover time of colored graphs. Throughout we only consider graphs that can be covered from every node. This property is analogous to the connectivity property for undirected graphs since without it there is no bound on the expected cover time.

We first summarize our results and then describe motivation and applications. We use the following notation in stating our results.

*Supported by NSF Grant CCR-9100886. Computer Science Department, University of Wisconsin, Madison, WI 53706.

†Supported by NSF Grant CCR88-13632. Computer Science Division, University of California, Berkeley, CA 94720.

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 obtained from G by removing all edges except those of color c the *underlying graph* of color c. We denote by A_c the $n\times n$ stochastic matrix whose $\{i,j\}$-th entry is the probability of reaching j from i in one step of the random walk when an edge of color c is followed.

We obtain the following bounds.

1. Exponential upper and lower bounds on the expected time to cover undirected graphs with two colors.

2. Double exponential upper and lower bounds on the expected time to cover undirected graphs with three or more colors.

3. Polynomial bounds on the expected cover time of colored graphs when the underlying graphs are aperiodic and have the same stationary distribution.

4. 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 $1/\mathrm{poly}(n)$.

We define the *expected cover time of G on a random sequence* to be the expected cover time of G on color sequence $C = C_1C_2C_3\ldots$, where the C_i are i.i.d. random variables that take on value j in $\{1,2,\ldots k\}$ with constant probability α_j. It is not hard to see that the expected cover time of any graph on a random sequence is at most exponential in the number of nodes of G. We prove the following bounds on the expected cover time of a graph on a random sequence.

5. An exponential lower bound on the expected time to cover colored graphs on a random sequence.

6. Polynomial bounds on the expected time to cover colored graphs on a random sequence, when all entries of the stationary distribution of $\sum_j \alpha_j A_j$ are at least $1/\mathrm{poly}(n)$.

As we will see, (undirected) colored graphs are a natural generalization of 2-colored directed graphs. Results on the expected cover time of 2-colored directed graphs were critical in determining the complexity of space bounded interactive proof systems [4], and that work motivated our definition of the expected cover time of a colored graph (in terms of the supremum). Undirected colored graphs are also a natural generalization of both directed and undirected graphs. Along with known polynomial bounds on the expected time to cover undirected graphs, our first two results give a complete characterization of cover times for colored graphs. Our definition of the expected cover time of a colored graph on a random sequence is a natural alternative definition of cover time, and our fifth result above, together with some straightforward observations, provides a complete characterization of the expected cover time of colored graphs on a random sequence.

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

Our results also have applications to the theory of time inhomogeneous Markov chains, defined by a sequence of stochastic matrices $P_1, P_2, P_3, \ldots$ A finite set of stochastic matrices defines a family of time inhomogeneous Markov chains; each chain of the family corresponds to an infinite sequence of matrices from the set. Such families of Markov chains arise in coding theory [8]. Our techniques can be used to bound the rate of convergence of such chains to their absorbing states and to bound the time needed for ergodicity properties to be achieved. In this way, our results introduce a complexity theory perspective to problems on time inhomogeneous Markov chains.

The rest of the paper is organized as follows. In Section 2 we present our exponential and double exponential bounds on the expected cover time of undirected colored graphs. In Section 3 we strengthen the exponential lower bound of Section 2 by showing it holds for a restricted class of graphs, even on a random sequence of colors. Finally, in Section 4 we describe a number of conditions under which a colored graph can be covered in polynomial expected time.

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 2-colored undirected graph is $2^{\Theta(poly(n))}$, whereas the expected cover time of an undirected graph with 3 or more colors is $2^{2^{\Theta(n)}}$. We first present upper bounds in Theorems 2.1 and 2.2 and then present the lower bounds in Theorems 2.3 and 2.4.

Before presenting the upper bounds we make the following definitions.

Let G be an undirected colored graph and let s and t be two nodes of G. We say that *t is reachable from s on the color sequence $C = C_1 \ldots C_k$*, if there is a sequence of nodes $s = v_0, v_1, \ldots, v_k = t$ such that G contains an edge of color C_i between v_{i-1} and v_i, for $i = 1, 2 \ldots k$. We call $v_0, v_1, \ldots, v_k$ a *path from s to t on C*.

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

Proof: Suppose G can be covered from node s. Fix a color sequence $C_1 C_2 C_3 \ldots$ We consider the random walk on this sequence from node s in intervals of $l = 2^n$ steps. Order the nodes $1, \ldots, n$. 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 $\geq 1/n^l$ 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 $\geq 1/n^l$ 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 $i+1$, that is, 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 $\geq 1/n^l$. This is because at each of the first l' steps of interval $i+1$, with probability $\geq 1/n$, 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 $j = 1, 2, \ldots l$, let S_j be the set of nodes reachable from s_i on the color sequence $C_{il+1}C_{il+2}\ldots C_{il+j}$. Since each set S_j is a subset of $\{1, \ldots, n\}$, by the pigeonhole principle $S_j = S_k$ for some $0 \leq j < k \leq l$.

Now, let C' be the color sequence $C_1C_2\ldots C_{il+j}(C_{il+j+1}\ldots C_{il+k})^\omega$. On this sequence, with probability > 0 node t is never reached from s. This is because with probability > 0, 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_l, $1 \leq l \leq k$. This contradicts our assumption that G can be covered from s. $\Box$

Later we'll show that this bound is tight for graphs with three or more colors. Undirected two-colored graphs, however, are coverable in expected time $2^{n^{O(1)}}$. We give a proof of this now.

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

Proof: Suppose G can be covered from node s. Fix a color sequence $C_1C_2C_3\ldots$, 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 $l = (4n-3)(n-1)$. Order the nodes $1, \ldots, n$. 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 $\geq 1/n^l$ 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 special case that the sequence $C_1C_2\ldots C_l$ is a prefix of $(BR)^\omega$ or

$(RB)^\omega$, and then extend that argument to arbitrary sequences. In fact, in this special case an interval of length $2n-1$ suffices.

Lemma 2.1 *Node t is reachable from s on a prefix of the sequences $(BR)^\omega$ or $(RB)^\omega$ of length $\leq 2n-1$.*

Proof: We prove the lemma for the sequence $(BR)^\omega$. The argument for the sequence $(RB)^\omega$ is analogous. Call a finite prefix C' of $(BR)^\omega$ *good* if t is reachable from s on C'. Since G can be covered from s, a good prefix exists. We show that there exists a good prefix of length $\leq 2n-1$.

Let C' be the shortest good prefix and let k be the length of C'. For contradiction assume that $k \geq 2n$. Let $s = w_0, w_1, \ldots, w_k = t$ be a path from s to t on C. Since k is $\geq 2n$, some node w appears in this path at least three times. So w appears twice in even numbered positions or twice in an odd numbered positions of the path. Without loss of generality, assume that w appears twice in even numbered positions, and let those positions be $2i$ and $2j$, with $i < j$. Let W be the sequence of nodes $s = w_0, w_1, w_2, \ldots, w_{2i-1}, w_{2j}, \ldots, w_k = t$. Then W is a path from s to t corresponding to a prefix of $(BR)^\omega$ of length $< k$, contradicting of our choice of C'. $\square$

We now generalize the argument to arbitrary sequences $C_1 \ldots C_l$. To do this, we relate arbitrary color sequences to prefixes of $(BR)^\omega$, 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.

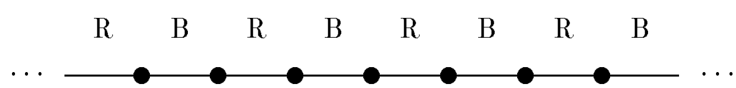

Figure 1: Line graph L

For clarity, we refer to the nodes of L as *points* in what follows, 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 equals the unique point reached on C'. The following simple lemma is the key to extending Lemma 2.1 to general sequences.

Lemma 2.2 *Suppose C is similar to C', where C' is a prefix of $(BR)^\omega$ or $(RB)^\omega$. 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: Intuitively, this lemma is true for the following reason. Suppose 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 just as are edges from p to q in L. On color sequence C, we construct a path from x to y in 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 $x = x'_0, x'_1, \ldots, x'_k = y$ be the path from x to y in G and let $p = p'_0, p'_1, \ldots, p'_k = q$ be the path from p to q in L, both on the sequence $C' = C'_1 C'_2 \ldots C'_k$. Let $p = p_0, p_1, \ldots, p_{l'} = q$ be the path from p to q in L on the color sequence $C = C_1 \ldots C_{l'}$. We construct a path $x = x_0, x_1, \ldots, x_{l'} = y$ in G on the color sequence $C_1 \ldots C_{l'}$.

This path is defined inductively as follows. We let $x_0 = x$. Suppose $0 < j \leq l'$ and that $x_1, \ldots, x_{j-1}$ are defined. Then x_j is defined as follows.

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

□

We now continue the proof that t is reachable from s on $C_1 \ldots C_{l'}$, for some $l' \leq l$. There are two cases. 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 either (i) $2n-1$ distinct points to the right of p or to the left of p are visited on the sequence $C_1 \ldots C_l$, or (ii) some point of L is visited n times on the sequence $C_1 \ldots C_l$. In the next two lemmas we show that in both cases t is reachable from s on $C_1 \ldots C_{l'}$, for some $l' \leq l$.

Lemma 2.3 *Suppose that $2n-1$ distinct points to the right of p (or to the left of p) are visited on the sequence $C_1 \ldots C_l$. 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 B.

From Lemma 2.1, on some prefix $C' = C'_1 \ldots C'_k$ of $(BR)^\omega$, where $k \leq 2n-1$, t is reachable from s in G. Let q be the point reachable from p on L on the color sequence $C'_1 \ldots C'_k$. 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. Then from Lemma 2.2, t is reachable from s on $C_1 \ldots C_{l'}$ as required. □

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 point q is visited at steps $j_1 < j_2 < \ldots < j_n \leq l$. Let S_i be the set of nodes reachable from s on the sequence $C_1 \ldots C_i$, $1 \leq i \leq l$. If t is contained in some $S_{l'}$, for $1 \leq l' \leq j_n - 1$, we are done since then t is reachable on the sequence $C_1 \ldots C_{l'}$. Suppose that t is not contained in any S_i, $1 \leq i \leq j_n - 1$. We show that t is contained in S_{j_n}, and hence t is reachable from s on $C_1 \ldots C_{l'}$, where $l' = j_n$.

To show that t is contained in S_{j_n}, we show that S_{j_i} is a proper subset of $S_{j_{i+1}}, 1 \leq i < n$. Then, since S_{j_1} contains at least 1 node, S_{j_n} must contain n nodes and hence must contain t.

The fact that $S_{j_i} \subseteq S_{j_{i+1}}$ follows from Lemma 2.2 because the color sequences $C_1 \ldots C_{j_i}$ and $C_1 \ldots C_{j_{i+1}}$ are similar. Now suppose that for some i, S_{j_i} is not a proper subset of $S_{j_{i+1}}$; that is, $S_{j_i} = S_{j_{i+1}}$. Then on the infinite sequence $C_1 C_2 \ldots C_{j_i}(C_{j_i+1} \ldots C_{j_{i+1}})^\omega$ t is never reached. This is because the only nodes reached on this sequence are those in the sets $S_1, \ldots, S_{j_{i+1}-1}$, and we are assuming that t is not contained in any of these sets. This contradicts the assumption that G can be covered from s. □

□

Theorem 2.3 shows that the bound of Theorem 2.2 is tight. The proof is based on the following lemma.

Lemma 2.5 *For every (uncolored) strongly-connected directed graph G there is a 2-colored undirected graph G' with the following properties:*

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. *the expected cover time of G' on $(RB)^\omega$ is the same (up to constant factors) as the cover time of G.*

Proof: Let G be a strongly-connected directed graph with vertex set $V = \{u_1, u_2, \ldots u_n\}$ and edge set E.

Let G' be a 2-colored undirected graph with vertex set $V' = \{v_1, v_2, \ldots, v_n, w_1, w_2, \ldots, w_n\}$ and edge set $E_R \cup E_B$. E_R and E_B are the sets of red and blue edges, respectively, and are defined as follows. $E_R = \{\{v_i, w_i\} | 1 \leq i \leq n\} \cup \{\{w_i, w_j\} | 1 \leq i \leq j \leq n\}$ and $E_B = \{\{w_i, v_j\} | (u_i, u_j) \in E\} \cup \{\{v_i, v_j\} | 1 \leq i \leq j \leq n\}$.

We now prove (2), that G' is covered from all its nodes. Suppose to the contrary that from some start node s, on some sequence of colors $C = C_1 C_2 C_3 \ldots$ a node v in G' is visited with probability < 1. Without loss of generality assume that $v \in \{v_1, \ldots, v_n\}$. (The case when $v \in \{w_1, \ldots, w_n\}$ is similar.) First, note that C must contain infinitely many B's. Otherwise, $C = C_1 \ldots C_i R^\omega$ and since (V', E_R) is a connected undirected graph, v is reached with probability 1 on this sequence.

Moreover, there must be some node $t \in \{v_1, \ldots v_n\}$ such that (i) on the color sequence $C_1 \ldots C_i$, there is a path from s to t which does not visit v, and (ii) v is not reachable from t on any prefix of $C_{i+1} C_{i+2} \cdots$. Because of the way G' is constructed, (ii) implies that no node in $\{v_1, \ldots, v_n\}$ is reachable from t on $C_{i+1} \ldots C_{i+j}$, where $j \geq 0$ and $C_{i+j+1} = B$. It follows that C must be of the form $C_1 \ldots C_i (RB)^\omega$.

But we claim that a random walk from t on $(RB)^\omega$ visits v with probability 1. This is because a random walk on this color sequence simulates a random walk of G starting at an arbitrary vertex. Because G is strongly-connected, this implies that the random walk covers G' from s, contradicting our assumption that v is visited with probability < 1. Hence G' can be covered from all of its nodes, completing the proof of (2).

This argument also proves (3), that on $(RB)^\omega$, the expected cover time of G' from a node v_i is the same (up to constant factors) as the expected cover time of G from node u_i. $\Box$

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

Proof: We can simply apply Lemma 2.5 to the strongly-connected directed graph with vertex set $\{u_1, u_2, \ldots u_n\}$ and edge set $\{(u_i, u_{i+1}) | 1 \leq i \leq n-1\} \cup \{(u_i, u_1) | 2 \leq i \leq n\}$. $\Box$

We can generalize this argument to show that the double exponential upper bound of Theorem 2.1 is tight for graphs with three or more colors. To do this we need the notion of a strongly-connected, directed 2-colored graph. A 2-colored directed graph G is *strongly-connected* if for every sequence of colors $C = C_1 C_2 C_3 \ldots$, $C_i \in \{R, B\}$, and every pair of vertices u and v, u is reachable from v on some prefix of C. Note that a strongly-connected graph can be covered from all its nodes.

Lemma 2.6 *For every strongly-connected 2-colored directed graph G there is a 3-colored undirected graph G' with the following properties:*

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 2-color sequence C, there exists a 3-color sequence C' such that the expected cover time of G' on C' is the same (up to constant factors) as the expected cover time of G on C.*

Proof: The construction here is similar to the one in Lemma 2.5. Let G be a strongly-connected 2-colored directed graph with vertex set $\{u_1, u_2, \ldots u_n\}$ and edge set $E_R \cup E_B$, where E_R and E_B are directed red and blue edges, respectively. Let G' be the following 3-colored undirected graph. The vertex set of G' is $\{v_1, v_2, \ldots v_n, w_1, w_2, \ldots w_n\}$ and the edge set is $E'_R \cup E'_B \cup E_Y$, where E'_R, E'_B, E_Y are sets of red, blue, and yellow edges, respectively, and are defined as follows. $E_Y = \{\{v_i, w_i\} | 1 \leq i \leq n\} \cup \{\{w_i, w_j\} | 1 \leq i \leq j \leq n\}$, $E'_R = \{\{w_i, v_j\} | (u_i, u_j) \in E_R\} \cup \{\{v_i, v_j\} | 1 \leq i \leq j \leq n\}$, and $E'_B = \{\{w_i, v_j\} | (u_i, u_j) \in E_B\} \cup \{\{v_i, v_j\} | 1 \leq i \leq j \leq n\}$.

The argument that G' can be covered from all its nodes is similar to that of Lemma 2.5. Note that if there is a sequence of colors on which a vertex $v \in \{v_1, \ldots, v_n\}$ is visited from some start node s with probability < 1, the sequence would have to be of the form

$C_1 \ldots C_i Y C'_1 Y C'_2 Y C'_3 Y \ldots$, where the $C'_j \in \{R, B\}$, and on the sequence $C_1 \ldots C_i$, a node t in the set $\{v_1, \ldots v_n\}$ is reachable. But a random walk from t on this subsequence would simulate a 2-color random walk of G. Because G is strongly-connected this implies that the random walk will cover G' with probability one.

For part (3), if $C = C_1 C_2 C_3 \ldots$, with $C_i \in \{R, B\}$, then let $C' = Y C_1 Y C_2 Y C_3 Y \ldots$ □

Theorem 2.4 *There are 3-colored undirected graphs that can be covered from all nodes and have expected cover time* $2^{2^{\Omega(n)}}$.

Proof: In [4] Condon and Lipton construct a family of strongly-connected directed 2-colored graphs with $O(n)$ nodes and expected cover time $2^{2^{\Omega(n)}}$. On a particular sequence of colors a random walk in the nth graph in the family simulates 2^n tosses of a fair coin and reaches an absorbing state only if all outcomes were heads. The result follows immediately by applying Lemma 2.6 to this graph. □

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 2-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 underlying graphs in a graph with self-loops are connected, 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 matrices.

Theorem 3.1 *There are undirected colored graphs with self loops which have expected cover time* $2^{\Theta(n)}$ *on the sequence* $(RB)^\omega$ *and, in fact, 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 the nodes $1, \ldots, n$ the *primary* nodes of the graph, and the nodes $1', \ldots n'$ the *secondary* nodes of the graph. 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 $i+1$.

The construction ensures that for $2 \leq i \leq n-1$, $p(i, i+1) = 1/2 - \epsilon$, for some constant $\epsilon > 0$, which 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. Suppose we define a 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.

$$T(i, i+1) = p(i, i+1) + (1 - p(i, i+1))(1 + T(i-1, i) + T(i, i+1)) \text{ and } T(1, 2) = 1$$

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

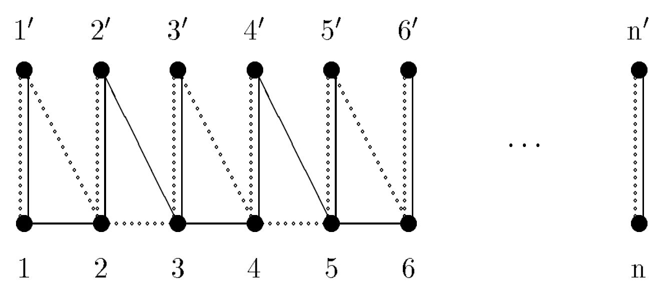

Figure 2: Exponential time graph with self-loops

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. $\square$

4 Polynomial Special Cases

In this section we identify several conditions under which undirected colored graphs are coverable 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 $(C_1 C_2 \dots C_l)^\omega$, 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 i-th 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 an undirected colored 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, we are taking a random walk in an aperiodic undirected graph). We will generalize this later to arbitrary sequences.

Let v_t be the n-vector whose i-th 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 j-th 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 [6], and Mihail [7] show that for some $t = \text{poly}(n)$, $\| \Delta_t \| \leq 1/\exp(n)$. The exact polynomial depends on the cutset expansion of the graph and is bounded above by n^3. The proof in [7] shows this by obtaining the 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 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 $C_1 C_2 C_3 \ldots$ is the color sequence, let v'_t be the probability vector for a random walk on $C_1 C_2 \ldots C_t$ 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 $t = n^3$, $\| \Delta'_t \| \leq 1/\exp(n)$. By definition, for all i, $\pi(i) \geq 1/n^2$, so $v'_t(i) \geq 1/cn^2$, c a constant > 0.

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 n^3 steps of a random walk. This analysis gives an $O(n^5 \log n)$ bound on the expected cover time. $\square$

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 undirected colored graphs are covered in polynomial expected time on sequences of the form $(C_1 C_2 \ldots C_l)^\omega$, 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 an undirected colored graph with n nodes which is connected in each color, and let $C_1 C_2 \ldots C_l$ 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 $(C_1 C_2 \ldots C_l)^\omega$ is $O(n^{l+2} p(n))$.

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

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

Let i and j be a pair of nodes in G_P such that $P_{i,j} > 0$. 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 $P_{i,j}$. Hence, the expected number of returns to i until the edge from i to j is traversed is $1/P_{i,j}$.

If $P_{i,j} = 1$, the expected time to traverse the edge from i to j is 1, and we are done. In what follows we assume that $P_{i,j} < 1$.

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 $T(i,i)/(1-P_{i,j})$. Hence, the expected time for the walk to traverse the edge from i to j is at most $T(i,i)/P_{i,j}(1-P_{i,j})$.

Since $P = A_{C_1}A_{C_2}\cdots A_{C_l}$ and each non-zero entry of the A_{C_i} is $\geq 1/n$, $P_{i,j} > 0$ implies that $P_{i,j} \geq 1/n^l$. Also note that $1 - P_{i,j} \geq 1/n^l$. Hence, $P_{i,j}(1-P_{i,j}) \geq (1/n^l)(1-1/n^l) \geq 1/2n^l$, and the expected time for the walk to traverse the edge from i to j is $\leq 2n^l T(i,i)$.

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 $\leq 2n^l p(n)$.

It follows that the expected time to cover G_P is $\leq 2n^{l+2}p(n)$.

□

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

Let G_P be a weighted, directed graph with probability transition matrix P. What the proof of Theorem 4.2 shows is that if P is irreducible and has all non-zero entries at least $1/\text{poly}(n)$, and all entries of the stationary distribution of P are at least $1/\text{poly}(n)$, the expected time to cover G_P is polynomial.

Using this idea again we get a similar result about the expected time to cover undirected colored graphs on random sequence of colors.

Theorem 4.3 *Let G be an undirected colored 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 time step color i is chosen with probability α_i, is polynomial.

Proof: The matrix $\sum_i \alpha_i A_i$ is irreducible and has all entries at least $1/\text{poly}(n)$. Having made this observation the rest of the proof is analogous to that of Theorem 4.2. $\square$

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

A more refined argument based on the techniques of Aleliunas et al. [1] and Göbel and Jagers [5] improves upon Theorem 4.2 in the following special case. We omit the proof here.

Theorem 4.4 *Let G be an undirected colored 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 Conclusions and Open Problems

We give tight bounds on the expected cover time of undirected colored graphs. We show that, in general, the expected cover time is exponential for two colors, and double exponential for three or more colors.

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 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.

We also give polynomial bounds on the expected time to cover colored graphs on repeated sequences and randomly chosen sequences when all entries of the associated stationary distributions are at least $1/\text{poly}(n)$. These results show that it is possible for the stationary distribution of a product or weighted average of matrices to contain exponentially small entries, even when all entries of the stationary distributions of the original matrices are at least $1/\text{poly}(n)$.

6 Acknowledgments

Richard Lipton first encouraged us to consider random walks in undirected colored graphs. We would like to thank Milena Mihail for helping us to better understand results on rapidly mixing Markov chains and their applicability to our work. Thanks to Prasoon Tiwari for many helpful discussions, and his contributions to Theorem 4.4. We also thank Dana Randall and Madhu Sudan for their comments on the writeup of this paper.

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), 1986, 83-96.

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

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

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

[6] 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.

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

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