Coalescing-Branching Random Walks on Graphs

Chinmoy Dutta [*]
Twitter
chinmoy@twitter.com

Gopal Pandurangan [†]
Nanyang Technological University
Brown University
gopalpandurangan@gmail.com

Rajmohan Rajaraman [‡]
Northeastern University
rraj@ccs.neu.edu

Scott Roche [§]
Northeastern University
str@ccs.neu.edu

ABSTRACT

We study a distributed randomized information propagation mechanism in networks we call the *coalescing-branching* random walk (cobra walk, for short). A cobra walk is a generalization of the well-studied "standard" random walk, and is useful in modeling and understanding the Susceptible-Infected-Susceptible (SIS)-type of epidemic processes in networks. It can also be helpful in performing light-weight information dissemination in resource-constrained networks. A cobra walk is parameterized by a *branching factor* k. The process starts from an arbitrary node, which is labeled *active* for step 1. (For instance, this could be a node that has a piece of data, rumor, or a virus.) In each step of a cobra walk, each active node chooses k random neighbors to become active for the next step ("branching"). A node is active for step $t+1$ only if it is chosen by an active node in step t ("coalescing"). This results in a stochastic process in the underlying network with properties that are quite different from both the standard random walk (which is equivalent to the cobra walk with branching factor 1) as well as other gossip-based rumor spreading mechanisms.

We focus on the *cover time* of the cobra walk, which is the number of steps for the walk to reach all the nodes, and derive almost-tight bounds for various graph classes. Our main technical result is an $O(\log^2 n)$ high probability bound for the cover time of cobra walks on expanders, if either the expansion factor or the branching factor is sufficiently large; we also obtain an $O(\log n)$ high probability bound for the *partial cover time*, which is the number of steps needed for the walk to reach at least a constant fraction of the nodes. We show that the cobra walk takes $O(n \log n)$ steps on any n-node tree for $k \geq 2$, and $\tilde{O}(n^{1/d})$ steps on a d-dimensional grid for $k \geq 2$, with high probability.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Stochastic processes; Probabilistic Algorithms; G.2.2 [**Discrete Mathematics: Graph Theory**]: Graph algorithms

General Terms

Algorithms, Theory

Keywords

Random Walks, Networks, Information Spreading, Cover Time, Epidemic Processes

1. INTRODUCTION

We study a distributed propagation mechanism in networks, called the *coalescing-branching random walk* (*cobra walk*, for short). A cobra walk is a variant of the standard random walk, and is parameterized by a *branching factor*, k. The process starts from an arbitrary node, which is initially labeled *active*. For instance, this could be a node that has a piece of data, rumor, or a virus. In a cobra walk, for each discrete time step, each active node chooses k random neighbors (sampled independently with replacement) to become active for the next step; this is the "branching" property, in which each node spawns multiple independent random walks. A node is active for step t if and only if it is chosen by an active node in step $t-1$; this is the "coalescing" property, i.e., if multiple walks meet at a node, they coalesce into one walk.

[*]Chinmoy Dutta was at Northeastern University when this work was done, and supported by NSF grant CCF-0845003 and a Microsoft grant to Ravi Sundaram.

[†]Supported in part by Nanyang Technological University grant M58110000, Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 2 grant MOE2010-T2-2-082, and MOE AcRF Tier 1 grant MOE2012-T1-001-094, and by a grant from the United States-Israel Binational Science Foundation (BSF).

[‡]Partially supported by NSF grants CCF-1216038 and CSR-1217812. Part of this work was done when Rajmohan Rajaraman was a Visiting Scientist at Google Research.

[§]Partially supported by NSF grant CSR-1217812 and by a grant from the United States-Israel Binational Science Foundation (BSF).

A cobra walk generalizes the standard random walk [35, 39], which is equivalent to a cobra walk with $k = 1$. Random walks on graphs have a wide variety of applications, including being fundamental primitives in distributed network algorithms for load balancing, routing, information propagation, gossip, and search [16, 17, 8, 44]. Being local and requiring little state information, random walks and their variants are especially well-suited for self-organizing dynamic networks such as Internet overlay, ad hoc wireless, and sensor networks [44]. As a propagation mechanism, one parameter of interest is the *cover time*, the expected time it takes to cover all the nodes in a network. Since the cover time of the standard random walk can be large — $\Theta(n^3)$ in the worst case, $\Theta(n \log n)$ even for expanders [35] — some recent studies have studied simple adaptations of random walks that can speed up cover time [1, 5, 18]. Our analysis of cobra walks continues this line of research, with the aim of studying a lightweight information dissemination process that has the potential to improve cover time significantly.

Our primary motivation for studying cobra walks is their close connection to SIS-type epidemic processes in networks. The SIS (standing for Susceptible Infected Susceptible) model (e.g., [20]) is widely used for capturing the spread of diseases in human contact networks or propagation of viruses in computer networks. Three basic properties of an SIS process are: (a) a node can infect one or more of its neighbors ("branching" property); (b) a node can be infected by one or more of its neighbors ("coalescence" property) and (c) an infected node can be cured and then become susceptible to infection at a later stage. Cobra walks satisfy all these properties, while standard random walks and other gossip-based propagation mechanisms violate one or more. Also, while there has been considerable work on the SIS model ([28, 43, 31, 20, 40, 19, 6]), it has been analytically hard to tackle basic *coverage* questions: (1) How long will it take for the epidemic to infect, say, a constant fraction of network? (2) Will every node be infected at some point, and how long will this take? Our analysis of cobra walks in certain special graph classes is a step toward a better understanding of such questions for SIS-type processes.

1.1 Our results and techniques

We derive near-tight bounds on the cover time of cobra walks on trees, grids, and expanders. These special graph classes arise in many distributed network applications, especially in the modeling and construction of peer-to-peer (P2P), overlay, ad hoc, and sensor networks. For example, expanders have been used for modeling and construction of P2P and overlay networks, grids and related graphs have been used as models for ad hoc and sensor networks, and spanning trees are often used as backbones for various information propagation tasks.

We begin with an observation that Matthew's Theorem [37, 35] for random walks extends to cobra walks; that is, the cover time of a cobra walk on an n-node graph is at most $\ln n$ times the maximum hitting time of a node. Hitting time is the expected time until a walk originating at $u \in V$ reaches $v \in V$ for the first time. For many graphs, this bound is also a tight bound. This enables us to focus on deriving bounds for the hitting time.

We face two technical challenges in our analysis. First, unlike in a standard random walk, cobra walks have multiple "active" nodes at any step, and in almost all graphs, it is difficult to characterize the distribution of the active nodes at any point of time. Second, the combination of the branching and coalescing properties introduces a non-trivial dependence among the active nodes, making it challenging to quantify the probability that a given node is made active during a given time period. Surprisingly, these challenges manifest even in tree networks. We present a result that gives tight bound on the cover time for trees, which we obtain by establishing a recurrence relation for the expected time taken for the cobra walk to cross an edge along a given path of the tree.

- For an arbitrary n-node tree, a cobra walk with $k \geq 2$ covers all nodes in $O(n \log n)$ steps with high probability (w.h.p., for short)[1] (Theorem 5 of Section 3.1).

For a matching lower bound, we note that the cover time of a cobra walk in a star graph is $\Omega(n \log n)$ w.h.p. We conjecture that the cover time for *any n-node graph* is $O(n \log n)$. By exploiting the regular structure of a grid, we establish improved and near-tight bounds for the cover time on d-dimensional grids.

- For a d-dimensional grid, we show that a cobra walk with $k \geq 2$ takes $\tilde{O}(n^{1/d})$ steps, w.h.p. (cf. Theorem 8 of Section 3.2).

Our main technical result is an analysis of cobra walks on expanders, which are graphs in which every set S of nodes of size at most half the number of vertices has at least $\alpha|S|$ neighbors for a constant α, which is referred to as the expansion factor.

- We show that for an n-node constant-degree expander, a cobra walk covers a constant fraction of nodes in $O(\log n)$ steps and all the nodes in $O(\log^2 n)$ steps w.h.p. assuming that either the branching factor or the expansion factor is sufficiently large (cf. Theorems 9 and 10 of Section 4).

Our analysis for expanders proceeds in two phases. We show that in the first phase, which consists of $O(\log n)$ steps, the branching process dominates resulting in an exponential growth in the number of active nodes until a constant fraction of nodes become active, with high probability. In the second phase, though a large fraction of the nodes continues to be active, dependencies caused by the coalescing property prevent us from treating the process as multiple independent random walks, analyzed in [2] (or even d-wise independent walks for a suitably large d). We overcome this hurdle by carefully analyzing these dependencies and bounding relevant conditional probabilities, and define a time-inhomogeneous Markov process that is stochastically dominated by the cobra walk in terms of coverage. We then use the notion of merging conductance and the machinery introduced in [38] to analyze time-inhomogeneous Markov chains, and establish an $O(\log n)$ bound w.h.p. on the maximum hiting time, leading to an $O(\log^2 n)$ bound on the cover time.

1.2 Related work and comparison

Branching and coalescing processes. There is a large body of work on branching processes (without coalescence)

[1] By the term "with high probability" (w.h.p., for short) we mean with probability $1 - 1/n^c$, for some constant $c > 0$.

on various discrete and non-discrete structures [33, 36, 4]. A study of coalescing random walks (without branching) was performed in [15] with applications to voter models. Others have looked at processes that incorporate branching and coalescing particle systems [3, 41]. However, these studies treat the particle systems as continuous-time systems, with branching, coalescing, and death rates on restricted-topology structures such as integer lattices. To the best of our knowledge, ours is the first work that studies random walks that branch and coalesce in discrete time and on various classes of non-regular finite graphs.

Random walks and parallel random walks. Feige [24, 23] showed that the cover time of a random walk on any undirected n-node connected graph is between $\Theta(n \log n)$ and $\Theta(n^3)$ with both the lower and upper bounds being achieved in certain graphs. With the rapidly increasing interest in information (rumor) spreading processes in large-scale networks and the gossiping paradigm (e.g., see [9] and the references therein), there have been a number of studies on speeding up the cover time of random walks on graphs. One of the earliest studies is due to Adler et al [1], who studied a process on the hypercube in which in each round a node is chosen uniformly at random and covered; if the chosen node was already covered, then an uncovered neighbor of the node is chosen uniformly at random and covered. For any d-regular graph, Dimitrov and Plaxton showed that a similar process achieves a cover time of $O(n + (n \log n)/d)$ [18]. For expander graphs, Berenbrink et al showed a simple variant of the standard random walk that achieves a linear (i.e., $O(n)$) cover time [5].

It is instructive to compare cobra walks with other mechanisms to speed up random walks as well as with gossip-based rumor spreading mechanisms. Perhaps the most related mechanism is that of parallel random walks which was first studied in [7] for the special case where the starting nodes are drawn from the stationary distribution, and in [2] for arbitrary starting nodes. Nearly-tight results on the speedup of cover time as a function of the number of parallel walks have been obtained by [22] for several graph classes including the cycle, d-dimensional meshes, hypercube, and expanders. (Also see [21] for results on mixing time.) Though cobra walks are similar to parallel random walks in the sense that at any step multiple nodes may be selecting random neighbors, there are significant differences between the two mechanisms. First the cover times of these walks are not comparable. For instance, while k parallel random walks may have a cover time of $\Omega(n^2/\log k)$ for any $k \in [1, n]$ [22], a 2-branching cobra walk on a line has a cover time of $O(n)$. Second, while the number of active nodes in k parallel random walks is always k, the number of active nodes in any k-branching cobra walk is continually changing and *may not even be monotonic*. Most importantly, the analysis of cover time of cobra walks needs to address several dependencies in the process by which the set of active nodes evolve; we use the machinery of time-inhomogenous Markov chains to obtain the cover time bound for bounded-degree expanders (see Section 4).

The works of [16, 17] presented distributed algorithms for performing a standard random walk in sublinear time, i.e., in time sublinear in the length of the walk. In particular, the algorithm of [17] performs a random walk of length ℓ in $\tilde{O}(\sqrt{\ell D})$ rounds w.h.p. on an undirected network, where D is the diameter of the network. However, this speed up comes with a drawback: the message complexity of the above faster algorithm is much worse compared to the naive sequential walk which takes only ℓ messages. In contrast, we note that the speedup in cover time given by a cobra walk over the standard random walk comes only at the cost of a slightly worse message complexity.

Gossip-based mechanisms. Gossip-based information propagation mechanisms have also been used for information (rumor) spreading in distributed networks. In the most typical rumor spreading models, gossip involves either a push step, in which nodes that are aware of a piece of information (being disseminated) pass it to random neighbors, or a pull step, in which nodes that are unaware of the information attempt to extract the information from one of their randomly chosen neighbors, or some combination of the two. In such models, the knowledgeable nodes or the ignorant nodes participate in the dissemination problem in *every* round (step) of the algorithm. The main parameter of interest in many of these analyses is the number of rounds needed till all the nodes in the network get to know the information.

The rumor spreading mechanism that is most closely related to cobra walks is the basic push protocol, in which in every step every informed node selects a random neighbor and pushes the information to the neighbor, thus making it informed. Feige et al. [25] show that the push process completes in every undirected graph in $O(n \log n)$ steps, with high probability. Since then, the push protocol and its variants have been extensively analyzed both for special graphs, as well as for general graphs in terms of their expansion properties (see e.g., [10, 11, 12, 30, 29, 27, 26]). Again, though cobra walks and push-based rumor spreading share the property that multiple nodes are active in a given step, the two mechanisms differ significantly. While the set of active nodes in rumor spreading is monotonically nondecreasing, this is not so in cobra walks, an aspect that makes the analysis challenging especially with regard to full coverage. Furthermore, the message complexity of the push protocol can be substantially different than that of cobra. A simple example is the star network, which the push protocol covers in $\Theta(n \log n)$ steps with a message complexity of $\Theta(n^2 \log n)$, while the 2-branching cobra walk has both cover time and message complexity $\Theta(n \log n)$. This can be extended to show similar results for star-based networks that have been proposed as models for Internet-scale networks [14].

1.3 Applications

As mentioned at the outset, cobra walks are closely related to the SIS model in epidemics, but they may be easier to analyze using tools from random walk and Markov chain analyses. While the persistence time and epidemic density of SIS-type epidemic models are well studied [28, 34, 43], to the best of our knowledge the time needed for a SIS-type process to affect a large fraction (or the whole) of the network has not been well-studied. Our results and analyses of cobra walks on more general networks can be useful in predicting the time taken for a real epidemic process following an SIS-type model to spread in a network.

Cobra walks can also serve as a lightweight information dissemination protocol in networks, similar to the push protocol. As pointed out earlier, in certain types of networks, the message complexity incurred by a cobra walk to cover a network can be smaller than that for the push protocol. This can be useful, especially in infrastructure-less anonymous networks, where nodes don't have unique identities and and may not even know the number of neighbors. In such networks, it is difficult to detect locally when coverage is completed[2]. If nodes have a good upper bound on n (the network size), however, then nodes can terminate the protocol after a number of steps equal to the estimated cover time. In such a scenario, message complexity is also an important performance criterion.

2. PRELIMINARIES

Let G be a connected graph with vertex set V and edge set E, and let $|V| = n$. We define a coalescing-branching (cobra) random walk on G with branching factor k starting at some arbitrary $v \in V$ as follows: At time $t = 0$ we place a pebble at v. Then in the next and every subsequent time step, every pebble in G clones itself $k - 1$ times (so that there are now k pebbles at each vertex that originally had a pebble). Each pebble independently selects a neighbor of its current vertex uniformly at random and moves to it. Once all pebbles make their one-hop moves, if two pebbles are at the same vertex they coalesce into a single pebble, and the next round begins. In a cobra-walk, a vertex receive a pebble an arbitrary number of times.

For a time step t of the process, let S_t be the **active set**, the set of all vertices of G that have a pebble. We will use two different definitions of the neighborhood of S_t: Let $N(S_t)$ be the **inclusive** neighborhood, the union of the set of neighbors of all vertices in S_t (which can include members of S_t itself). Let $\Gamma(S_t)$ be the **non-inclusive neighborhood**, which is the union of the set of neighbors of all vertices of S_t such that $S_t \cap \Gamma(S_t) = \emptyset$.

Let the expected **maximum hitting time** h_{max} of a cobra-walk on G be defined as the $\max_{u,v \in V} \mathbb{E}[h_{u,v}]$ where $h_{u,v}$ is the time it takes a cobra-walk starting at vertex u to first reach v with at least one pebble.

We are interested in two different notions of cover time, the time until all vertices of G have been visited by a cobra-walk at least once. Let τ_v be the minimum time t such that, for a cobra-walk starting from v, $\forall u \in V - v$, $u \in S_t$ for some $t \leq \tau_v$ which may depend on u. Then we define the **cover time** of a cobra-walk on G to be $\max_{v \in V} \tau_v$. We define the **expected cover time** to be $\max_{v \in V} \mathbb{E}[\tau_v]$. Note that in the literature for simple random walks, cover time usually refers to the expected cover times. In this paper we will show high-probability bounds on the cover time.

In Section 6 we will be proving results for cobra-walks on expanders. In this paper, we will use a spectral definition for expanders and then use Tanner's theorem to translate that to neighborhood and cut-based notions of expanders.

Definition 1. *An* ϵ***-expander*** *graph is a d-regular graph whose adjacency matrix has eigenvalues α_i such that $|\alpha_i| \leq \epsilon d$ for $i \geq 2$.*

We also want to define the notion of an ϵ-approximation:

Definition 2. *G is an* ϵ***-approximation*** *for a graph H if $(1-\epsilon)H \preccurlyeq G \preccurlyeq (1+\epsilon)H$, where $H \preccurlyeq G$ if for all x, $x^T L_H x \leq x^T L_G x$, where L_G and L_H are the Laplacians of G and H, respectively.*

[2]In networks with identities and knowledge of neighbors, a node can locally stop sending messages when all neighbors have the rumor. This reduces the overall message complexity until cover time.

Finally, we will rely on the neighborhood expansion of a set S on G, where we define $N(S)$ as the inclusive neighborhood. For this we will use Tanner's theorem [42], which gives us a lower bound on the size of the neighborhood of S for sufficiently strong expanders.

Theorem 3. *Let G be a d-regular graph that ϵ- approximates $\frac{d}{n}K_n$. Then for all $S \subseteq V$ with $|S| = \delta n$, $|N(S)| \geq \frac{|S|}{\epsilon^2(1-\delta)+\delta}$.*

3. COVER TIME FOR TREES AND GRIDS

A useful tool in bounding the cover time for simple random walks is Matthew's Theorem [37, 35], which bounds the expected cover time of a graph by the maximum expected hitting time $h_{u,v}$ between any two nodes u and v times the harmonic number H_n. Here we show that this result can be extended to cobra walks. The full proof can be found in the full version of the paper, but the key idea is that we map the cobra walk on G to a simple walk on much larger graph derived from G and then show that this satisfies the conditions for the proof of Matthew's Theorem on simple walks.

Theorem 4. ***Matthew's Theorem for Cobra Walks*** *Let G be a connected graph on n nodes. Let w be a cobra walk on G starting at an arbitrary node. Then the cover-time of w on G, $C(G)$, is bounded from above by $h_{max} \ln n$ in expectation and by $O(h_{max} \ln n)$ with high probability.*

Matthew's theorem for cobra walks is used in proving the cover time for trees and grids.

3.1 Trees

Theorem 5. *For any tree, the cover time of a cobra walk starting from any node is $O(n \ln \ln n)$ w.h.p.*

We will prove our main result by calculating the maximum hitting time of a cobra walk on a tree T and then applying Matthew's theorem. Cobra walks on trees are especially tractable because they follow two nice properties. Since a tree has a unique path between any two nodes, we only need keep track of the pebble closes to the target. In addition, the fact that there is one simple path between any two nodes limits the number of collisions we need to keep track of, a property which is not true for general graphs and makes cobra walk harder to analyze on them. For this section, we fix the branching factor $k = 2$. For $k > 2$ but still constant, the cover time would not be asymptotically better.

The general idea behind the proof is as follows. We take the longest path w.r.t. hitting time in the tree. Along each node in this path, except for the first and last, there will be a subtree rooted at that node. If a cobra walk's closest pebble to the endpoint is at node l, the walk from this point can either advance with at least one pebble, or it can not advance by either backtracking along the path, going down the subtree rooted at l, or both. We show via a stochastic dominance argument that a biased random walk from l, whose transition probabilities are tuned to be identical to cobra walk's, will next advance to $l + 1$ in a time that is dominated primarily by the size of the subtree at l. This is done by analyzing the return times in the non-advancement scenarios listed above. Thus summing up over the entire walk, the hitting time is dominated by a linear function of the size of the entire tree.

In Lemma 6 we bound the return time of a cobra walk to a root of the tree.

LEMMA 6. *Let T be a tree of size M. Pick a root, r, and let r have d children. Then a cobra walk on T starting at r will have a return time to r of $O(4M/d)$.*

PROOF. To show that the Lemma holds for a cobra walk, we will actually show that it holds for a simple random walk with transition probabilities modified to resemble those of a cobra walk. For this simple random walk, we start at r and in the first step pick one of the children of r, r'. Let $(d'+1)$ be the degree of r'. Then we define transition probabilities as follows: p is the probability of returning to r in the next step, and q is the probability of continuing down the tree. They are given as: $p = \left(1 - \left(\frac{d'}{(d'+1)}\right)^2\right)$, $q = \left(\frac{d'}{(d'+1)}\right)^2$, $\frac{p}{q} = \frac{(d')^2}{(2d'+1)}$. Note that these are the exact same probabilities that a cobra walk at node r' would have for sending (not sending) at least one (any) pebbles back to the root.

The rest of the proof follows by mathematical induction. Consider a tree T that has only two levels. Starting from r, the return time, 2, is constant, the relationship holds. For the inductive case, assume that the hypothesis holds. Then:

$$\begin{aligned} r(T) &\leq 1 + \sum_{r' \in N(r)} p(r') h_{r',r} \leq 1 + \frac{1}{d} \sum_{r' \in N(r)} h_{r',r} \\ &\leq 1 + \frac{1}{d} \sum_{r' \in N(r)} \left(1 + \frac{d'^2}{2d'+1} c \frac{|T'|}{d'}\right) \\ &\leq 2 + \frac{c|T|}{2d} \end{aligned}$$

Setting $c = 4$ gives us the result of the lemma for the biased random walk, and it is easy to see that by stochastic dominance this holds also for the cobra walk. $\square$

Finally, we show a key lemma for the hitting time of a single step of a path along a tree.

LEMMA 7. *Fix a path in a tree T made up of nodes $1, \ldots, l, (l+1), \ldots, t$. Then, the expected time it takes for a cobra walk starting at node l to get to $l+1$ with at least one pebble is given by:*

$$h_{l,(l+1)} = \frac{5}{4} + \frac{12}{5} \sum_{i=l}^{2} \left(\frac{1}{5}\right)^{l-i} |T_i| \tag{1}$$

where T_l is the induced subtree formed by taking node l, its neighbors not on the path being traversed, and all of their descendants.

Informally, we prove that the one-step hitting time is bounded by above by the worst case scenario that either both pebbles go back along the path or down the subtree rooted at l and establish a simple recurrence relation.

PROOF. Vertex l is viewed through the context of having one edge to the node $l-1$, one edge to node l, and d edges to some other nodes. Thus it can be viewed as the root of a tree, and T_l as the induced subgraph of l and all nodes reached through its d_l not-on-path children. We will need the following probabilities:

- Probability of a pebble going from l to $l+1 = p = \left(1 - \left(\frac{(d_l+1)}{(d_l+2)}\right)^2\right)$
- Probability of a pebble not going from l to $l+1 = 1-p = q$.
- Probability of a cobra walk sending both pebbles from l to $l-1$ conditioned on it *not* sending any pebbles from l to $l+1 = q_l^{'} = \left(\frac{1}{(d_l+1)^2}\right)$
- Probability of a cobra walk sending at least one pebble to the subtree T_l conditioned on its not sending any pebbles to $l+1 = q_l^{''} = \left(\frac{(d_l)}{(d_l+1)}\right)^2 + 2\left(\frac{d_l}{(d_l+1)^2}\right) = \frac{d_l^2 + 2d_l}{(d_l+1)^2}$

Note that, conditioned on a pebble not advancing to node $l+1$, we actually have three disjoint events: (A) Both pebbles go to $l-1$, (B) one pebble goes to $l-1$ and one pebble goes into subtree T_l, and (C) both pebbles go into T_l. We define an alternate event B', which is the event that one pebble goes down T_l and nothing else happens (thus, it is not technically in the space of cobra walk actions). If we let R be the time until first return of the cobra walk to l conditioned on no pebble going to $l+1$, we wish to show that $E[R|B] \leq E[R|B']$ and that $E[R|C] \leq E[R|B']$. What is the relationship between B and B'? Consider two random variables, X and Y, and let X be the time until first return of a pebble that travels from l to $l-1$, Y be the time until first return of a pebble that travels into T_l. Then $R|B$ is just another random variable, $U = \min(X, Y)$. Since $U \leq Y$ over the entire space, $E[U] \leq E[Y]$, and clearly $R|B'$ is equivalent to Y. Thus $E[R|B] \leq E[R|B']$ It is also easy to see that $E[R|B'] \geq E[R|C]$. Thus by the law of total expectation we have:

$$\begin{aligned} E[R] &= E[R|A]\Pr(A) + E[R|B]\Pr(B) + E[R|C]\Pr(C) \\ &\leq E[R|A]\Pr(A) + (\Pr(B) + \Pr(C))E[R|B'] \\ &= E[R|A]\Pr(A) + E[R|B'](1 - \Pr(A)) \end{aligned}$$

Then the hitting time can be expressed as:

$$\begin{aligned} h_{l,l+1} &\leq p + q(E[R] + h_{l,l+1}) \\ \Rightarrow (1-q)h_{l,l+1} &\leq p + q(E[R]) \\ \Rightarrow h_{l,l+1} &\leq 1 + \frac{q}{p}(q_l^{'}(1 + h_{l-1,l}) + q_l^{''} r(T_l)) \end{aligned}$$

Note that $q/p = \frac{(d_l+1)^2}{(2d_l+3)}$. Since $r(T_l) \leq 4|T_l|/d_l$ by Lemma 6, we continue with: $h_{l,l+1} \leq 1 + \frac{(d_l+1)^2}{(2d_l+3)} \frac{1}{(d_l+1)^2}(1 + h_{l-1,l}) + \frac{(d_l+1)^2}{(2d_l+3)} \frac{(d_l^2+2d_l)}{(d_l+1)^2} \frac{4|T_l|}{d_l} \leq 1 + \frac{1}{5}(1 + h_{l-1,l}) + \frac{12}{5}|T_l|$ w.h.p.

If we expand the relation, we get: $h_{l,l+1} \leq \sum_{i=0}^{l} \left(\frac{1}{5}\right)^i + \frac{12}{5}\left(|T_l| + \left(\frac{1}{5}\right)|T_{l-1}| + \left(\frac{1}{5}\right)^2 |T_{l-2}| + \cdots + \left(\frac{1}{5}\right)^{l-2} |T_2|\right)$, and thus $h_{l,l+1} \leq \frac{5}{4} + \frac{12}{5} \sum_{i=l}^{2} \left(\frac{1}{5}\right)^{l-i} |T_i|$ $\square$

We are finally ready to prove our main results for the tree, Theorem 5, that the cobra walk cover time of an arbitrary tree occurs in $O(n \ln n)$ steps.

PROOF. By Matthew's Theorem for cobra walks, $C(G) \leq (\ln n + o(1))h_{max}$. We just need to prove that h_{max} occurs in linear time.

Let P be the path for which $h_{u,v}$ is maximized, and let the path consist of the sequence of nodes $1, 2, \ldots, t$. As in the proof of the single-step hitting time, we note that for all but the first and last nodes on P, there is a subtree T_l of size $|T_l|$ rooted at each nodes. Because $h_{1,t} \leq h_{1,2} + h_{2,3} + \ldots h_{t-1,t}$ we obtain the desired result from Lemma 7 as follows:

$$\begin{aligned} h_{1,t} &\leq \frac{5}{4}t + \frac{12}{5}\sum_{j=2}^{t-1}\left[|T_j|\sum_{i=0}^{\infty}\left(\frac{1}{5}\right)^i\right] \\ &\leq \frac{5}{4}t + \frac{12}{5}\frac{5}{4}\sum_{j=2}^{t-1}|T_j| \leq 4n. \end{aligned}$$

□

We note that for the line network, we can improve the bound we obtain for trees and show that the cover time of a cobra walk is $O(n)$ w.h.p.

3.2 Grids

For a d-dimensional grid, we show the following theorem whose proof is in the full version.

THEOREM 8. *Let G be a finite d-dimensional grid for some constant d, without wrap-around edges. Then the cover time of a cobra walk on G is $\tilde{O}(n^{1/d})$ w.h.p for branching factor $k = 2$.*

Here we present a sketch of the proof, which can be found in the long version of the paper.

Consider a cobra-walk which starts at the origin of the lattice $(0, \ldots, 0)$. In each step of the cobra-walk, we define the following process to determine which pebble we are tracking as we move towards the target vertex $(n^{1/d}, \ldots, n^{1/d})$. We focus on one dimension at a time. W.l.o.g. consider the first dimension. We refer to a $+1$ motion as movement in the "right" direction (towards the first coordinate of the target) and -1 as a movement away from the target in the first dimension. From the current node, two pebbles pick neighbors uniformly at random. We follow a pebble that makes most progress in the direction of the target along the first dimension. That is, if one of the pebbles goes to the neighbor in the $+1$ direction, we track that pebble and move to the node it selected. If both pebbles pick the -1 neighbor, we move to that neighbor. Otherwise, we randomly pick one of the nodes that were selected by the pebbles and note there was a movement of 0 in the first dimension. Projecting this process onto the first dimension, we have a biased random walk (towards the target) in that dimension. After $O(n^{1/d})$ steps, we can use a large deviation bound (e.g. Theorems 2.8 and 2.8 in [13]) to show that with constant probability, we reach the target's first coordinate value to within $n^{1/2d}$ distance.

What is happening in the other coordinates? It is easy to see that projecting our walk described above on any other dimension creates an unbiased random walk over that dimension. Hence, after $O(n^{1/d})$ steps, with constant probability, we are within $n^{1/2d}$ steps of the position we started in.

After approaching the target in the first dimension, we repeat the process in the other dimensions. Each dimension requires $O(n^{1/d})$ steps. At the end of the first phase, with constant probability, for each coordinate, we are at most $n^{1/2d}$ steps away from the target. We keep doing this for $\log\log n$ phases. At the end of phase i we are a distance at most $n^{\frac{1}{d2^{i+1}}}$ from the target with probability p^i for some constant $0 < p < 1$. After $\log\log n$ phases the walk will be within $O(1)$ distance (in each dimension) from the target with probability $1/\text{polylog}(n)$.

The last $O(1)$ steps to the target can be made by the walk by taking $O(1)$ "advance" steps, which will happen with at least a constant probability. Thus, in $O(n^{1/d})$ steps we reach the target with probability $1/\text{polylog}(n)$.

In expectation, we will reach the target after $O(n^{1/d}\text{polylog}(n))$ steps. Applying Matthew's bound yields the result of the lemma.

4. ANALYSIS FOR EXPANDERS

For expander graphs, we are able to prove a high probability cover time result of $O(\log^2 n)$. We break the proof up into two phases. In the first phase we show that a cobra walk starting from any node will reach a constant fraction of the nodes in logarithmic time w.h.p. In the second phase, we create a process which stochastically dominates the cobra walk and show that this new process, will cover the entire rest of the graph again in polylogarithmic time w.h.p.

The main result of this section can be stated in the following two theorems, which when taken together imply that w.h.p. ϵ-expander G will be covered in $O(\log^2 n)$ time.

THEOREM 9. *Let G be any ϵ-expander with ϵ, δ not depending on n (number of nodes in G), with $\delta < \frac{16}{30d^2}$, and ϵ, a sufficiently small constant such that*

$$\frac{1}{\epsilon^2(1-\delta)+\delta} > \frac{d(de^{-k}+(k-1)) - \frac{k^2}{2}}{d(e^{-k}+(k-1)) - \frac{k^2}{2}}, \tag{2}$$

then in time $O(\log n)$, w.h.p. a cobra walk on G with branching factor k, will attain an active set of size δn.

We note that the condition in the above theorem is satisfied if either ϵ is sufficiently small, or k is sufficiently large. For instance when $k = 2$, the above condition holds for strong expanders, such as the Ramanujan graphs, which have $\epsilon \leq 2\sqrt{d-1}/d$, and random d-regular graphs, for d sufficiently large.

THEOREM 10. *Let G be as above, and let W be a cobra walk on G that at time T has reached an active set of size δn. Then w.h.p in an additional $O(\log^2 n)$ steps every node of G will have visited by W at least once.*

To prove Theorem 9 we prove that active sets up to a constant fraction of V are growing at each step by a factor greater than one. The proof can be found in the full version.

LEMMA 11. *Let G be any ϵ-expander with ϵ, δ satisfying the conditions of Theorem 9. Then for any time $t \geq 0$, the cobra walk on G with active set S_t such that $|S_t| \leq \delta n$ satisfies $\mathbb{E}[|S_{t+1}|] \geq (1+\nu)|S_t|$ for some constant $\nu > 0$.*

Next, we use a standard martingale argument to show that the number of nodes in S_t is concentrated around its expectation. The proof of Lemma 12 can also be found in the full version of the paper.

LEMMA 12. *For a cobra walk on a d-regular ϵ-expander that satisfies the conditions in Lemma 11, at any time t*

$$\Pr\left[|S_{t+1}| - \mathbb{E}[|S_{t+1}|] \le -\tau |S_t|\right] \le e^{-\frac{\tau^2 |S_t|}{2k}} \quad (3)$$

Finally, using the bound of Lemma 12 we show that with high probability we will cover at least δn of the nodes of G with a cobra walk in logarithmic time by showing that the active set for some $t = O(\log n)$ is of size at least δn.

LEMMA 13. *For a cobra walk on d-regular, ϵ-expander G, there exists a time T such that $T = O(\log n)$ and $|S_T| \ge \delta n$.*

The key to proving Lemma 13 is to view a cobra-walk on G as a Markov process over a different state space consisting of all of the possible sizes of the active set. In this interpretation, all configurations of pebbles in a cobra-walk in which i nodes are active are equivalent. The goal is to show that this new Markov process will reach a state corresponding to an active set of size δn quickly w.h.p. To prove this, we first show that it is dominated by a restricted Markov chain over the same state space in which any negative growth in the size of the active set is replaced with a transition to the initial state (in which only one node is active). We then in turn show that the restricted walk is dominated by an even more restricted walk in which the probability of negative growth is higher than in the first restricted walk, bounded from below from a constant, and no longer dependent on the size of the current state. We then show that the goal of the lemma is achieved even in this walk by relating the process to a negative binomial random variable.

PROOF. We view a cobra-walk on G as a random walk W over the state space consisting of all of the possible sizes of the active set: $S(W) = \{1, \dots, n\}$. We then define a Markov process M_1 that stochastically dominates W: Let $\tau = \nu/2$, where ν is the expected growth factor of the active set as shown in Lemma 11. The states of M_1, $S(M_1)$ are the same as W's, but the transitions between states differ. Each $i \in S(W)$ can have out-arcs to many different states, but the corresponding $i \in S(M_1)$ has only two transitions. With probability $p_i = 1 - e^{-\frac{\nu^2 i}{8k}}$ transition to state $(1+\nu/2)i$, and with probability $1 - p_i$ transition to state 1. Note that p_i is derived from Lemma 12.

In M_1, each transition probability is still a function of the current state i, and as mentioned above we would like to eliminate this dependence. Thus, define M_2 as a random walk over the same state space. However, we will deal only with a subset of $S(M_2)$: the states: $(1 + \nu/2)^i C$ for $i \in \mathbb{Z}$ and a suitably large constant C. We then have the following transitions for each state in the chain (which will begin once it hits C). Setting $r = \nu^2/8k$, at state $(1 + \nu/2)^i C$: 1) Transition to state $(1 + \nu/2)^{i+1} C$ with probability $p'_i = 1 - e^{-rC(1+\frac{i\nu}{2})}$ 2) Transition to state C with probability $1 - p'_i$. This Markov chain oscillates between failure (going to C) and growing by a factor of $1 + \nu/2$. Note that to get success (i.e., reaching a state of at least δn), we need $\Omega(\log n)$ growing transitions.

The probability that in a walk on this state space that we "fail" and go back to C before hitting δn is bounded by 1/2, since $\sum_{i=0}^{\infty} e^{-rC(1+i\frac{\nu}{2})} \le e^{-rC} \sum_{i=0}^{\infty} e^{irC\frac{\nu}{2}} = \frac{e^{-rC}}{1-e^{-rC\frac{\nu}{2}}} \le \frac{1}{2}$, provided that C is sufficiently large as a function of r (which is itself only a function of the branching factor and the constant ν).

Consider each block of steps that end in a failure (meaning we return to C). Then clearly w.h.p. after $b \log n$ trials, for some constant b, we will have a trial that ends in success (i.e., reaching an active set of size δn nodes). In these $b \log n$ trials, there are exactly that many returns to C. However, looking across all trials that end in failure, there are also only a total of $O(\log n)$ steps that are successful (i.e., involve a growth rather than shrinkage). To see why this is true, note that the probability of a failure after a string of growth steps goes down supralinearly with each step, so that if we know we are in a failing trial it is very likely that we fail after only a few steps. Thus, there cannot be too many successes before each failure. Indeed, the probability that we fail at step i within a trial can be bounded. Thus $\Pr[\text{Failure at step i} \mid \text{eventual failure}]$

$$\begin{aligned}
&= \frac{\Pr[\text{Failure at step i}]}{\Pr[\text{Eventual failure}]} \\
&= \frac{e^{-rC(1+i\nu/2)}}{\sum_{i=1}^{\infty}\left(\prod_{j=1}^{l-1}(1-e^{-rC(1+j\nu/2)}\right)e^{-rC(1+l\nu/2)}} \\
&\ge \frac{1}{\sum_{i=1}^{\infty} e^{-irC\nu/2}} \ge 1 - e^{-rC\nu/2}
\end{aligned}$$

and thus the probability of advancing is no more than $e^{-rC\nu/2}$, also a quantity that does not depend on i. This is a negative binomial random variable with distribution $w(k, p)$, the number of coin flips needed to obtain k heads with heads probability p. Identifying heads with a failure (i.e. returning to C) and tails with making a growth transition, we have a random variable $w(k, p)$, the number of coin flips needed for k failures with probability of failure $p = 1 - e^{-rC\nu/2}$. It is well known that $\Pr[w(k, p) \le m] = \Pr[B(m, p) \ge k]$, where $B(m, p)$ is the binomial random variable counting the number of heads within m p-biased coin flips. Thus, $\Pr[w(k, p) > m] = \Pr[B(m, p) < k]$. Setting $k = a \log n$ and $m = b \log n$, we have, $\Pr[B(m, p) \le \mathbb{E}[B(m, p)] - t] = \Pr[B(m, p) < pm - t] \le e^{\frac{-2t^2}{m}}$. We let $k = pm - t$, and solving for t we get $t = (pb - a) \log n$. This gives us

$$\Pr[B(m, p) < k)] \le \frac{1}{n^{\frac{(pb-a)^2}{b}}},$$

establishing there are at most $O(\log n)$ success within $O(\log n)$ trials ending in failure. Via stochastic dominance this bound holds for our original cobra walk process.

□

Once the active set has reached size $\Omega(n)$, we need a different method to show that the cobra-walk achieves full coverage in $O(\log^2 n)$ time. We can not simply pick a random pebble and restart the cobra-walk from this point $O(\log n)$ times because we know nothing about the distribution of the δn pebbles after restart, and the restarting method would require the pebbles to be i.i.d. uniform across the nodes of G. As a result, we are unable to establish a straightforward bound on h_{max} and invoke Matthew's Theorem.

Hence, we develop a different process, which we will call W_{alt}, that is stochastically dominated by the cobra walk. In W_{alt}, no more branching or coalescing occurs, and we also modify the transition probabilities of the pebbles on a node-by-node basis, depending on the number of pebbles at a node.

Definition 14. *For any time t and any collection of S pebbles on V (there can be more than 1 pebble at a node), define $W_{alt}(t+1)$ as follows. Let $A \subseteq V$ be the set of all nodes with 1 pebble at time t. Let $B \subseteq V$ be the set of all nodes with exactly 2 pebbles, and let C be the set of all nodes with more than 2 pebbles. Then, (a) for every $v \in A$, the pebble at v uniformly at random selects a node in $N(v)$ and moves to it; (b) for every $v \in B$, each pebble at v uniformly at random selects its own node in $N(v)$ and moves to it; (c) for every $v \in C$, arbitrarily order the pebbles at v, the first two pebbles then pick a neighbor to hop to uniformly at random. The remaining pebbles then pick with probability $1/2$ one of the two neighbors already selected and move to that node.*

If at time t a node during process W_{alt} has two or more pebbles, at each time step it behaves identically to a node running a cobra walk. On the other hand, if there is only one pebble at node running W_{alt} it acts like a simple random walk. Thus the number of active nodes at the next time step in W_{alt} is a (possibly proper) subset of the nodes with pebbles if the graph were running the cobra walk instead. Since this will be true at every time step, W_{alt} stochastically dominates the cobra walk w.r.t cover time τ of G, and it will be enough to prove the following:

Theorem 15. *Let G be a bounded-degree d-regular ϵ-expander graph, with ϵ sufficiently high to satisfy the conditions in Lemma 11. Let there be δn pebbles distributed arbitrarily over V, with at most one pebble per node. Let $\delta < \dfrac{16}{30d^2}$. Let λ be the second-largest eigenvalue of the adjacency matrix of G. From our ϵ-expander definition, $\lambda \leq \epsilon d$. For every ϵ, there is a constant ϵ' that is the node expansion constant of G. Furthermore, let constant $\gamma = \frac{\epsilon'}{\epsilon^2(1-\delta)+\delta}$, and let $s = \frac{5\log n + 6\log d + \log 9}{-\log\left(1-\frac{1}{2}\left(\frac{\gamma}{64d^{10}}\right)^2\right)}$. Starting from this configuration, the cover time of W_{alt} on G is $O(\log^2 n)$, with high probability.*

Proof. Our proof relies on showing that each node in G has a constant probability of being visited by at least one pebble during an epoch of W_{alt} lasting $\Theta(\log n)$ time. Once this has been established, all nodes of G will be covered w.h.p. after $O(\log n)$ epochs lasting $\Theta(\log n)$ steps each.

Define E_i to be the event that pebble i covers an arbitrary node v in s steps. We want to prove that the probability that v is covered by at least one pebble, $\Pr\left[\bigcup_i E_i\right]$, is constant. Using a second-order inclusion-exclusion approximation:

$$\begin{aligned}\Pr\left[\bigcup_i E_i\right] &\geq \sum_i \Pr\left[E_i\right] - \sum_{i\neq j}\Pr\left[E_i \cap E_j\right]\\ &= \sum_i \Pr\left[E_i\right] - \sum_{i\neq j}\Pr\left[E_i\right]\Pr\left[E_j|E_i\right].\end{aligned}$$

As a marginal probability, $\Pr\left[E_i\right]$ can be viewed as the probability that the random walk of pebble i hits v at time s. Thus, we only need to look at the elements of zA^i, where A is the stochastic matrix of the simple random walk on G and z is a vector with $z(l) = 1$ for the l, the position of pebble i at the beginning of the epoch and 0 in all other positions. In [2] it is proved in Lemma 4.8 that each coordinate of $A^{s'}z$ differs from $1/n$ by at most $\frac{1}{2n}$ for $s' = \frac{\ln 2n}{\ln \epsilon}$. Since $s > s'$, this hold for our case as well. Thus $\Pr\left[E_i = 1\right] \geq \frac{1}{2n}$.

Next we establish an upper bound for $\Pr\left[E_j|E_i\right]$. Due to the conditioning on the walk of pebble i, we can't use the transition matrix A^i, but we would like to do something similar. The transition matrix governing the walk of pebble j conditioned on a fixed walk of pebble i can be characterized at each step by transition matrix $P_{l(i,t)}$, where $l(i,t)$ is the location of pebble i at time t, can be described as follows. For every row k of $P_{l(i,t)}$ s.t. $k \neq l(i,t)$ we have an exactly copy of the k^{th} row of A, the transition matrix of an independent random walk on G. When $k = l(i,t)$ this represents the walk of j when pebbles i and j are co-located at node k. To establish an upper bound, we assume the worst case, that j is ordered as the 3rd or higher pebble at k. Let τ be the neighbor of node k chosen by pebble i. Then $P[k,\tau] = 1/2 + 1/2d$, and for all other positions of row k where A is non-zero, the corresponding position in $P = 1/2d$. These represent the transition probabilities according to W_{alt} as described earlier.

From an initial probability distribution z chosen over $V(G)$, the probability of pebble j being at node v conditioned on the walk of pebble i is the v^{th} component of $z\prod_{t=1}^{s} P_{l(i,t)}$. In Lemma 16 we show that the largest component of $z\prod_{t=1}^{s} P_{l(i,t)}$ is no more than $\frac{5d^2}{2n}$. With this result, we then have:

$$\begin{aligned}\Pr\left[\bigcup E_i\right] &\geq \sum_i \Pr\left[E_i\right] - \frac{1}{2}\sum_{i\neq j}\Pr\left[E_i\right]\Pr\left[E_j|E_i\right]\\ &\geq \delta n\frac{1}{2n} - \frac{1}{2}\binom{\delta n}{2}\frac{3}{2n}\frac{5d^2}{2n}\\ &\geq \frac{\delta}{2} - \frac{15}{16}\delta^2 d^2\end{aligned}$$

which will be a constant for the sufficiently small δ (depending only on d) given in the statement of the Theorem. ☐

Lemma 16. *Let G, γ, ϵ', and s be as stated in Theorem 15. Let i and j be two pebbles walking according to the rules of W_{alt} on G. Fix the walk of i, and let $\{P_{l(i,t)}\}$ be the sequence of perturbed transition matrices for the walk of pebble j depending on i. Then starting i from an arbitrary node, after s steps, the probability that j is at any node is at most $5d^2/2n$.*

Proof. The proof of this lemma relies heavily on Theorem 3.2 in [32], which we review and state here. Let P be an irreducible, ergodic Markov process for which reversibility and strong aperiodicity are not required. Consider the weighted transition from state i to j, $w_{ij} = \pi_i p_{ij}$, where π_i is the stationary distribution of i and p_{ij} is the transition probability from i to j of P. For $A \subset V$, we define the **merging conductance** of set A as

$$\Phi^*_P(A) = \frac{\sum_{j_1\in A}\sum_{j_2\in V-A}\sum_i \frac{w_{j_1 i}w_{j_2 i}}{\pi_i}}{\sum_{i\in A}\pi_i} \tag{4}$$

The merging conductance of graph G is thus

$$\Phi^*_P(G) = \min_{A\subset S:\sum_{i\in A}\pi_i \leq \frac{1}{2}} \Phi^*_P(A)$$

Intuitively, the merging conductance can be viewed as a measure of the flow coming into all nodes from both A and

$V-A$ for some set A. The higher the merging conductance of a graph, the more well connected it is and evenly distributed the flow is. If we define $\|\vec{x}(t)\| = \sum \frac{(p_i(t)-\pi_i)^2}{\pi_i}$ to be a measure of the distance of a distribution $\vec{p}$ over V from the stationary distribution of P, then [38] gives us the following theorem, which indicates that for a graph ğwith merging conductance bounded away from zero, convergence to the stationary distribution occurs in logarithmic time.

THEOREM 17 ([38, THEOREM 3.2]). *For an arbitrary initial distribution $\vec{x}(0)$ over V*

$$\|\vec{x}(t)\| \leq (1-\frac{1}{2}(\Phi_P^*)^2)^t \|\vec{x}(0)\|$$

We also need the following lemma for bounds on the maximum and minimum of the stationary distribution of the conditional walk of pebble j.

LEMMA 18. *For the walk of pebble j as described on W_{alt} for a suitable d-regular ϵ-expander G, conditioned on the walk of pebble i. the stationary distribution of the walk of j has bounds $\pi_{min} \geq \frac{1}{2nd^2}$ and $\pi_{max} \leq \frac{2d^2}{n}$.*

Next we establish a lower bound for the number of terms in the sum in the numerator of Equation 4. Let A be the set for which $\Phi_P^*(G)$ is minimized. Furthermore, since G is an ϵ-expander, we also know that its cobra walk expansion is a constant ϵ' and depends only on ϵ. We would like to calculate the number of nodes in G that have at least one neighbor in A and at least one neighbor in $V-A$. First, we lower-bound the size of the set of nodes with at least one edge to A. This set is just $N(A)$, the inclusive neighborhood of A, which from Tanner's theorem can be bounded from below by $\frac{|A|}{\epsilon^2(1-\delta)+\delta}$. Of the node in $N(A)$, we also need to bound the number that also have at least one edge to $V-A$. However, this is just the non-inclusive neighborhood of $N(A)$, $\Gamma(N(A))$, and we can use the node expansion of G to show that $|\Gamma(N(A))| \geq \frac{\epsilon'|A|}{\epsilon^2(1-\delta)+\delta}$. Thus we get:

$$\begin{aligned}
\Phi_P^*(G) &\leq \frac{\epsilon'|A|}{\epsilon^2(1-\delta)+\delta}\frac{\frac{\pi_{min}^2(1/2d)^2}{\pi_{max}}}{|A|\pi_{max}} \\
&\leq \frac{\epsilon'}{\epsilon^2(1-\delta)+\delta}\left(\frac{1}{2d^2}\right)^2\left(\frac{1}{2d}\right)^2\frac{1}{n^2}\left(\frac{n}{2d^2}\right)^2 \\
&\leq \frac{\epsilon'}{\epsilon^2(1-\delta)+\delta}\frac{1}{64d^{10}}
\end{aligned}$$

Letting $\gamma = \frac{\epsilon'}{\epsilon^2(1-\delta)+\delta}$, we note that the expression above is a constant as long as $d, \epsilon, \delta, \epsilon'$ are constants, which will be true in a d-regular ϵ-expander.

Starting from a distribution $\vec{x}(0)$ whose norm $\|\vec{x}(0)\|$ will be maximized when the walk is started from node s.t. $\pi_i = \pi_{min}$, we have:

$$\begin{aligned}
\|\vec{x}(0)\| &\leq \frac{(1-\pi_{min})^2}{\pi_{min}} + (n-1)\frac{(\pi_{max})^2}{\pi_{min}} \\
&\leq 2d^2n + (n-1)\left(\frac{2d^2}{n}\right)^2(2d^2n) \\
&\leq 2d^2n + 8d^6 < 9d^6n
\end{aligned}$$

for $d > 1$. Finally, we want to show that $\|\vec{x}(s)\| < \frac{1}{n^4}$. With this, it is clear to see that the maximum difference $|p_i(t)-\pi_i| < \frac{1}{n^2}$ which implies that the maximum probability $\Pr[E_j|E_i] < \frac{2d^2}{n} + \frac{1}{n^2} < \frac{5d^2}{2n}$ as required in Theorem 15. To do this we need to show that $\left(1-\frac{1}{2}\left(\frac{\gamma}{64d^{10}}\right)^2\right)^s \leq \frac{1}{9d^9n^5}$, which will be true for the set value of s in the definition of the Theorem.

A final note: because $\Phi_P^*(G) \leq \frac{\gamma}{64d^{10}}$ for every matrix $P_{l(i,t)}$, we can apply Theorem 17 in the exponentiation even though each matrix is different. □

5. CONCLUSION

We studied a generalization of the random walk, namely the cobra walk, and analyzed its cover time for trees, grids, and expander graphs. The cobra walk is a natural random process, with potential applications to epidemics and gossip-based information spreading. We plan to explore further the connections between cobra walks and the SIS model, and pursue their practical implications. From a theoretical standpoint, there are several interesting open problems regarding cobra walks that remain to be solved. First is to obtain a tight bound for the cover time of cobra walks on expanders. Our upper bound is $O(\log^2 n)$, while the diameter $\Omega(\log n)$ is a basic lower bound. Another pressing open problem is to determine the worst-case bound on the cover time of cobra walks on general graphs. It will also be interesting to establish and compare the message complexity of cobra walk with the standard random walk and other gossip-based rumor spreading processes.

Acknowledgments

The authors would like to thank the reviewers and are especially grateful to the reviewer who suggested an improved result and proof for the cover time of grids.

6. REFERENCES

[1] M. Adler, E. Halperin, R. M. Karp, and V. V. Vazirani. A stochastic process on the hypercube with applications to peer-to-peer networks. In *STOC*, pages 575–584, 2003.

[2] N. Alon, C. Avin, M. Koucký, G. Kozma, Z. Lotker, and M. R. Tuttle. Many random walks are faster than one. In *SPAA*, pages 119–128, 2008.

[3] S. R. Arthreya and J. M. Swart. Branching-coalescing particle systems. *Probability theory and related fields*, 131(3):376–414, 2005.

[4] I. Benjamini and S. Müller. On the trace of branching random walks. *arXiv preprint arXiv:1002.2781*, 2010.

[5] P. Berenbrink, C. Cooper, R. Elsässer, T. Radzik, and T. Sauerwald. Speeding up random walks with neighborhood exploration. In *SODA*, pages 1422–1435, 2010.

[6] N. Berger, C. Borgs, J. T. Chayes, and A. Saberi. On the spread of viruses on the internet. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 301–310. Society for Industrial and Applied Mathematics, 2005.

[7] A. Broder. Generating random spanning trees. In *FOCS*, 1989.

[8] M. Bui, T. Bernard, D. Sohier, and A. Bui. Random walks in distributed computing: A survey. In *IICS*, pages 1–14, 2004.

[9] J.-Y. Chen and G. Pandurangan. Almost-optimal gossip-based aggregate computation. *SIAM J. Comput.*, 41(3):455–483, 2012.

[10] F. Chierichetti and S. L. andAlessandro Panconesi. Almost tight bounds for rumour spreading with conductance. In *STOC*, pages 399–408, 2010.

[11] F. Chierichetti, S. Lattanzi, and A. Panconesi. Rumour spreading and graph conductance. In *SODA*, pages 1657–1663, 2010.

[12] F. Chierichetti, S. Lattanzi, and A. Panconesi. Rumor spreading in social networks. *Theoretical Computer Science*, 412(24):2602–2610, 2011.

[13] F. Chung, L. Lu, C. B. of the Mathematical Sciences, and N. S. F. (U.S.). *Complex graphs and networks*. Number no. 107 in CBMS Regional Conference Ser. in Mathematics Series. American Mathematical Society, 2006.

[14] F. Comellas and S. Gago. A star based model for the eigenvalue power law of internet graphs. *Phys. A*, 351:680–686, 2005.

[15] C. Cooper, R. Elsässer, H. Ono, and T. Radzik. Coalescing random walks and voting on graphs. In *Proceedings of the 2012 ACM symposium on Principles of distributed computing*, pages 47–56. ACM, 2012.

[16] A. Das Sarma, D. Nanongkai, and G. Pandurangan. Fast distributed random walks. In *PODC*, pages 161–170, 2009.

[17] A. Das Sarma, D. Nanongkai, G. Pandurangan, and P. Tetali. Efficient distributed random walks with applications. In *PODC*, 2010.

[18] N. B. Dimitrov and C. G. Plaxton. Optimal cover time for a graph-based coupon collector process. In *ICALP*, pages 702–716, 2005.

[19] M. Draief and A. Ganesh. A random walk model for infection on graphs: spread of epidemics & rumours with mobile agents. *Discrete Event Dynamic Systems*, 21(1):41–61, 2011.

[20] R. Durrett. Some features of the spread of epidemics and information on a random graph. *Proceedings of the National Academy of Sciences*, 107(10):4491–4498, 2010.

[21] K. Efremenko and O. Reingold. How well do random walks parallelize? In *APPROX-RANDOM*, pages 476–489, 2009.

[22] R. Elsässer and T. Sauerwald. Tight bounds for the cover time of multiple random walks. In *ICALP (1)*, pages 415–426, 2009.

[23] U. Feige. A tight lower bound on the cover time for random walks on graphs. *Random Struct. Algorithms*, 6(4):433–438, 1995.

[24] U. Feige. A tight upper bound on the cover time for random walks on graphs. *Random Struct. Algorithms*, 6(1):51–54, 1995.

[25] U. Feige, D. Peleg, P. Raghavan, and E. Upfal. Randomized broadcast in networks. *Random Structures an Algorithms*, 1(4):447–460, 1990.

[26] N. Fountoulakis and K. Panagiotou. Rumor spreading on random regular graphs and expanders. In *APPROX-RANDOM*, pages 560–573, 2010.

[27] N. Fountoulakis, K. Panagiotou, and T. Sauerwald. Ultra-fast rumor spreading in social networks. In *SODA*, pages 1642–1660, 2012.

[28] A. Ganesh, L. Massoulie, and D. Towsley. The effect of network topology on the spread of epidemics. In *INFOCOM*, volume 2, pages 1455 – 1466, march 2005.

[29] Giakkoupis. Tight bounds for rumor spreading in graphs of a given conductance. In *STACS*, pages 57–68, 2011.

[30] G. Giakkoupis and T. Sauerwald. Rumor spreading and vertex expansion. In *SODA*, pages 1623–1641, 2012.

[31] O. Givan, N. Schwartz, A. Cygelberg, and L. Stone. Predicting epidemic thresholds on complex networks: Limitations of mean-field approaches. *Journal of Theoretical Biology*, 288:21–28, 2011.

[32] C. Gkantsidis, M. Mihail, and A. Saberi. Throughput and congestion in power-law graphs. In *SIGMETRICS*, pages 148–159, 2003.

[33] T. E. Harris. *The theory of branching processes*. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer-Verlag, Berlin, 1963.

[34] D. A. Kessler. Epidemic Size in the SIS Model of Endemic Infec- Tions. *ArXiv e-prints*, Sept. 2007.

[35] L. Lovász. Random walks on graphs: a survey. In *Combinatorics, Paul Erdös is Eighty*, pages 1–46. 1993.

[36] N. Madras and R. Schinazi. Branching random walks on trees. *Stochastic Processes and their Applications*, 42(2):255 – 267, 1992.

[37] P. Matthews. Covering problems for brownian motion on spheres. *The Annals of Probability*, pages 189–199, 1988.

[38] M. Mihail. Conductance and convergence of markov chains-a combinatorial treatment of expanders. In *Foundations of Computer Science, 1989., 30th Annual Symposium on*, pages 526–531. IEEE, 1989.

[39] M. Mitzenmacher and E. Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2004.

[40] R. Parshani, S. Carmi, and S. Havlin. Epidemic threshold for the susceptible-infectious-susceptible model on random networks. *Physical review letters*, 104(25):258701, 2010.

[41] R. Sun and J. M. Swart. The brownian net. *The Annals of Probability*, 36(3):1153–1208, 2008.

[42] R. M. Tanner. Explicit concentrators from generalized n-gons. *SIAM Journal on Algebraic Discrete Methods*, 5(3):287–293, 1984.

[43] P. Van Mieghem. The n-intertwined SIS epidemic network model. *Computing*, 93:147–169, 2011.

[44] M. Zhong and K. Shen. Random walk based node sampling in self-organizing networks. *Operating Systems Review*, 40(3):49–55, 2006.