Improved Cover Time Bounds for the Coalescing-Branching Random Walk on Graphs

Colin Cooper
Dept Informatics, King's College
London
London WC2R 2LS, U.K.
colin.cooper@kcl.ac.uk

Tomasz Radzik
Dept Informatics, King's College
London
London WC2R 2LS, U.K.
tomasz.radzik@kcl.ac.uk

Nicolás Rivera
Dept Informatics, King's College
London
London WC2R 2LS, U.K.
nicolas.rivera@kcl.ac.uk

ABSTRACT

We present improved bounds on the cover time of the *coalescing-branching random walk* process COBRA. The COBRA process, introduced in [Dutta et al., SPAA 2013], can be viewed as spreading a single item of information throughout an undirected graph in synchronised rounds. In each round, each vertex which has received the information in the previous round (possibly simultaneously from more than one neighbour and possibly not for the first time), 'pushes' the information to b randomly selected neighbours. The COBRA process is typically studied for integer branching rates $b \geq 2$ (with the case $b = 1$ corresponding to a random walk). The aim of the process is to propagate the information quickly, but with a limited number of transmissions per vertex per round.

The cover time of COBRA is defined as the expected number of rounds until each vertex has received the information at least once. Our main results are a bound of $O(m + (d_{\max})^2 \log n) = O(n^2 \log n)$ on the COBRA cover time for an arbitrary connected graph with n vertices, m edges and the maximum vertex degree $d_{\max}$, and a bound of $O((r^2 + r/(1-\lambda)) \log n)$ for r-regular connected graphs with the second eigenvalue λ. Our bounds improve the $O(n^{11/4} \log n)$ and $O((r^4/\phi^2) \log^2 n)$ bounds shown in [Mitzenmacher et al., SPAA 2016], where ϕ is the conductance of the graph, and complement the $O((1/(1 - \lambda))^3 \log n)$ bound shown in [Cooper et al., PODC 2016]. We obtain our bounds by analysing the process called *Biased Infection with Persistent Source* (BIPS), which was introduced in [Cooper et al., PODC 2016] as a dual process for COBRA .

CCS CONCEPTS

• **Theory of computation** → **Distributed algorithms**; **Random walks and Markov chains**; • **Mathematics of computing** → *Graph theory*;

KEYWORDS

Random processes on graphs; cover time; COBRA process; epidemic processes

1 INTRODUCTION

Dutta et al. [5, 6] studied the following *coalescing-branching* random walk process for propagating information on a connected graph with n vertices and m edges. At the start of a round each vertex containing information 'pushes' this information to b randomly selected neighbours. It then stops passing the information until it receives the information again. At the end of a round if a vertex receives information from two or more vertices, then the information coalesces into one. Thus it does not help if a vertex receives the same information from more than one neighbour. The continuous act of coalescing and branching gives the name COBRA to this process.

A COBRA process can be modelled as a particle process. At the start of each round, each existing particle divides into b particles (the branching factor). These particles then move independently to random neighbours. At the end of each round any particles which meet at a vertex coalesce to form a single particle.

The aim of the COBRA process is to rapidly propagate information to all vertices but to limit the number of transmissions per vertex per round and without requiring that vertices store information for longer than one round. In the special case that $b = 1$, the COBRA process is a simple random walk, which achieves a low transmission rate but does not satisfy the fast propagation condition.

The main quantity of interest in information propagation processes is the time taken to inform (or visit) all vertices. By analogy with a random walk, this is referred to as the *cover time*. The w.h.p.[1] cover time results for the COBRA process obtained in [5, 6] for the case $b = 2$ include the following. (i) For the complete graph K_n all vertices are visited in $O(\log n)$ rounds. (ii) For regular constant-degree expanders, the cover time is $O(\log^2 n)$. (iii) For the D-dimensional grid, the cover time is $\widetilde{O}(n^{1/D})$, where $\widetilde{O}(.)$ indicates the presence of a poly-log n term. Improved bounds were shown later in [8]: an $O((r^4/\phi^2) \log^2 n)$ bound for r-regular graphs with conductance ϕ, an $O(D^2 n^{1/D})$ bound for D-dimensional grids, and an $O(n^{11/4} \log n)$ bound for general graphs.

Let Diam(G) denote the diameter of a graph G. Then max$\{\log_2 n,$ Diam(G)$\}$ is a lower bound on the number of rounds needed for the COBRA process with branching factor $b = 2$ to inform every vertex. This is the best possible, since the number of visited vertices at most doubles in each round. By comparison with the complete graph, and considering this lower bound, it might seem that the cover time of an r-regular expander by the COBRA process with

---

[1] *With high probability* (w.h.p.) means with probability at least $1 - n^{-c}$, for some positive constant c.

branching factor $b = 2$ should be $O(\log n)$ for any degree r between 3 and $n-1$. This is indeed the case as proven in [4]: the cover time is $O((1/(1-\lambda))^3 \log n)$ for r-regular graphs with the second eigenvalue λ, for any $3 \le r \le n - 1$.

Our contributions. In this paper we show two new bounds on the cover time of the COBRA process for branching factor $b = 2$. For arbitrary connected graphs, we improve the $O(n^{11/4} \log n)$ bound given in [8] to $O(m + (d_{\max})^2 \log n) = O(n^2 \log n)$, where $d_{\max}$ is the maximum degree of a vertex. For r-regular connected graphs, we show a bound of $O\left((r^2 + r/(1-\lambda)) \log n\right)$, which improves the $O((1/(1-\lambda))^3 \log n)$ bound given in [4] for the case when $1 - \lambda = o(1/\sqrt{r})$. Both bounds require that $1-\lambda > C\sqrt{(\log n)/n}$, for a suitably large constant C. Since $1 - \lambda \ge \phi^2/2$, our new bound for regular graphs improves also the $O((r^4/\phi^2) \log^2 n)$ bound given in [8]. As an example, consider the hypercube with $n = 2^d$ vertices, which has degree $r = \log n$ and both the conductance ϕ and the *eigenvalue gap* $1 - \lambda$ are $\Theta(1/\log n)$. The bounds presented in [8], [4] and in this paper give, respectively, the following cover-time bounds for the hypercube: $O(\log^8 n)$, $O(\log^4 n)$ and $O(\log^3 n)$.

We proceed with the formal definition of the COBRA process and the statement of our main results. Consider a graph $G = (V, E)$, an integer $b \ge 1$ and a subset of vertices $C \subseteq V$. The COBRA process with starting set C and branching factor b is the set process $(C_t)_{t \ge 0}$ with $C_0 = C$ and C_{t+1} generated as follows. Each vertex $v \in C_t$ independently chooses b neighbours uniformly at random with replacement and all the chosen vertices belong to C_{t+1}. For $C_0 = \{u\}$, let $\mathbf{cover}(u) = \min\{T : \bigcup_{t=0}^{T} C_t = V\}$ be the number of rounds needed for the COBRA process to visit all vertices of the graph starting from vertex u; and let $\text{COVER}(u) = \mathbf{E}(\mathbf{cover}(u))$. By analogy with the cover time of a random walk, which uses the worst case starting vertex, we let $\text{COVER}(G) = \max_{u \in V} \text{COVER}(u)$ be the cover time of the COBRA process on graph G. We derive asymptotic upper bounds on $\mathbf{cover}(u)$, which hold w.h.p. for each vertex $u \in V$ (Theorems 1.1 and 1.2). By the argument of 'restarting' the process from the current state in the low probability event that the graph has not been covered by the claimed bound, we conclude that the same asymptotic bounds apply also to the expectation $\text{COVER}(u)$, so to $\text{COVER}(G)$ as well.

THEOREM 1.1. *Let G be a connected graph with n vertices, m edges and the maximum vertex degree $d_{\max}$. For the COBRA process with branching factor $b = 2$, w.h.p. for each $u \in V$, cover(u) is*

$$O(m + (d_{\max})^2 \log n). \tag{1}$$

For a connected r-regular graph G with adjacency matrix $A(G)$, let $P = A(G)/r$ denote the transition matrix of the random-walk on G. Let $\lambda_1, \lambda_2, ..., \lambda_n$ be the eigenvalues of matrix P ordered in a non-increasing sequence. Thus $\lambda_1 = 1$ and $\lambda_n \ge -1$. Let $\lambda = \lambda(G) = \max_{i=2,...,n} |\lambda_i|$ be the second largest eigenvalue of (in absolute values). Our second result gives a bound on the cover time of COBRA for regular graphs in terms of the eigenvalue gap $1 - \lambda$ and the vertex degree r. This bound assumes $1 - \lambda > 0$, which holds if and only if the graph is connected and not bipartite. We note that the same bound can be derived for bipartite connected graphs, if we consider the 'lazy' COBRA process, which allows each vertex to also select itself with probability $1/2$.

THEOREM 1.2. *Let G be a connected r-regular n-vertex graph with $1 - \lambda > C\sqrt{(\log n)/n}$, for some suitably large constant C. For the COBRA process with branching factor $b = 2$, w.h.p. for each $u \in V$, cover(u) is*

$$O\left(\left(\frac{r}{1-\lambda} + r^2\right) \cdot \log n\right). \tag{2}$$

The COBRA process is a type of multiple random walk processes, so it is tempting to try to analyse it using techniques developed for such processes. Previous work on multiple random walks includes [1–3, 7], where cover times were analysed for various classes of graphs, assuming that the random walks are independent. The analyses of the COBRA process given in Dutta et al. [5, 6] and Mitzenmacher et al. [8] use a number of tools from multiple random walks, but applicability of those tools turns out to be limited because the random walks in COBRA are highly dependent. An alternative approach was proposed by Cooper et al. [4], who introduced and analysed a related epidemic process BIPS and showed that it is a dual (in some sense) of the COBRA process under time reversal.

Biased Infection with Persistent Source (BIPS). For a graph $G = (V, E)$, an integer $b \ge 1$ and a vertex v in G, which acts as the 'persistent' source of an infection, we consider the set process $(A_t)_{t \ge 0}$ defined by $A_0 = \{v\}$ and the following rule for generating A_{t+1} from A_t. Given A_t, each vertex $u \in V$, other than v, independently and uniformly with replacement selects b neighbours and becomes a member of A_{t+1} if and only if at least one of the selected neighbours is in A_t. Additionally, $v \in A_t$ for all $t \ge 0$. We call A_t the infected set at time t. Observe the source v is always infected, while other vertices can keep changing their status between infected and not infected.

For a subset $S \subseteq V$, let $\text{Infect}(S)$ denote the random set infected from S in one round: each vertex $u \in V$ selects independently uniformly with replacement two neighbours and becomes a member of $\text{Infect}(S)$, if and only if, at least one of the two selected neighbours is in S. Thus the BIPS process starts with $A_0 = \{v\}$ and $A_t = \text{Infect}(A_{t-1}) \cup \{v\}$, for each round $t \ge 1$.

The BIPS process is a discrete epidemic process of the Susceptible-Infected-Susceptible (SIS) type. The dynamics of such a process specifies how vertices get infected from their neighbours and how they lose infection (turn back into the susceptible state). In the BIPS process, the vertices (other than the source v) refresh their infected state at each round by contacting b randomly chosen neighbours. The presence of a persistent (or corrupted) source means that all vertices of the underlying graph eventually become infected. Our main reason for considering this particular BIPS process is that it is dual to the COBRA process. The BIPS process may however be also of independent interest since in the context of epidemics, certain viruses exhibit the property that a particular host can become persistently infected.

To avoid confusion between the BIPS process $(A_t)_{t \ge 0}$ and the COBRA process $(C_t)_{t \ge 0}$, we use the notation $\mathbf{P}(\cdot)$ for probabilities in the BIPS process, and $\widehat{\mathbf{P}}(\cdot)$ in the COBRA process. Let $\text{Hit}(w) = \min\{t : w \in C_t\}$, that is, the round in the COBRA process when vertex w gets a particle for the first time. We also write $\text{Hit}_C(w)$ and $\text{Hit}_u(w)$, if we want to indicate that $C_0 = C$ and $C_0 = \{u\}$, respectively. Our main results for COBRA follow from the duality between COBRA and BIPS introduced in [4] and expressed in the following theorem.

THEOREM 1.3. [4] *Let G be a connected graph and consider the COBRA and BIPS processes on G with the same parameter $b \geq 1$. For each $v \in V$ (the persistent source in BIPS), any set $\emptyset \neq C \subseteq V$ (the initial set with particles in COBRA) and $T \geq 0$, we have*

$$\widehat{\mathbf{P}}(\text{Hit}(v) > T | C_0 = C) = \mathbf{P}(C \cap A_T = \emptyset | A_0 = \{v\}).$$

This theorem says that the probability that a vertex v is not hit by round T in COBRA starting with particles at each vertex of set C is equal to the probability that none of the vertices in C is infected *at* round T in BIPS with the persistent source v. This theorem is formally proven in [4], so here we give only the underlying idea. Fix the sets of neighbours which the vertices select in rounds $1, 2, \ldots, T$: $\omega(u, t) \subseteq \mathcal{N}(u)$, for $u \in V$, $1 \leq t \leq T$, where $\mathcal{N}(u)$ denotes the set of neighbours of vertex u. Run the COBRA process for T rounds using these selections of neighbours, that is, if vertex u receives a particle in round $t - 1$, then it sends in round t one particle to each neighbour in $\omega(u, t)$. Run also the BIPS process using the same selections of neighbours but in the reverse time order, that is, the sets of neighbours $\omega(v, t)$, $v \in V$, are used in rounds $T + 1 - t$. It turns out that in the COBRA process vertex v is visited within these T rounds, if and only if, in the BIPS process at least one vertex in C is infected at round T. Crucially, the probability that the COBRA process selects exactly these sets $\omega(u, t)$ is equal to the probability that the BIPS process selects exactly the same sets in the reverse time order.

We define the infection time $\mathbf{infec}(v)$ as the first time when all vertices are infected by the BIPS process with the persistent source v. Proofs of the following two theorems about the BIPS process are the main new technical contribution of this paper.

THEOREM 1.4. *Let G be a connected graph with n vertices, m edges and the maximum vertex degree $d_{\max}$. For every $v \in V$, the infection time $\mathbf{infec}(v)$ of the BIPS process with $b = 2$ satisfies bound (1) with probability at least $1 - O(1/n^3)$.*

THEOREM 1.5. *Let G be a connected n-vertex r-regular graph with $1 - \lambda > C\sqrt{(\log n)/n}$, for some suitably large constant C. For every $v \in V$, the infection time $\mathbf{infec}(v)$ of the BIPS process with $b = 2$ satisfies bound (2) with probability at least $1 - O(1/n^3)$.*

Theorems 1.1 and 1.2 follow from Theorems 1.4 and 1.5, respectively, and from Theorem 1.3. For any two vertices $u, v \in V$ and any $T \geq 0$, applying Theorem 1.3 with $C = \{u\}$ gives

$$\begin{aligned}
\widehat{\mathbf{P}}(\text{Hit}_u(v) > T) &= \mathbf{P}(u \notin A_T | A_0 = v) \\
&\leq \mathbf{P}(A_T \neq V | A_0 = v) \\
&= \mathbf{P}(\mathbf{infec}(v) > T).
\end{aligned}$$

Theorem 1.4 says that there is a constant $c > 0$ such that for $T = c(m + (d_{\max})^2 \log n)$, $\mathbf{P}(\mathbf{infec}(v) > T) = O(1/n^3)$, implying $\widehat{\mathbf{P}}(\text{Hit}_u(v) > T) = O(1/n^3)$. We have $\mathbf{cover}(u) > T$, if and only if there is a vertex v such that $\text{Hit}_u(v) > T$. Thus, using the union bound, we conclude that $\mathbf{cover}(u)$ is greater than T with probability $O(1/n^2)$ and Theorem 1.1 follows. To see that the expected value of $\mathbf{cover}(u)$ is $O(T)$, consider restarting the COBRA process after T rounds from any vertex in C_T, if the graph has not yet been covered. We obtain Theorem 1.2 from the corresponding Theorem 1.5 in an analogous way.

COBRA process with branching factor less than 2.
Theorems 1.1 and 1.2 are proved for a COBRA process with branching factor $b = 2$. However, it seems natural to ask if cover times of the same order can be obtained with less branching. Clearly $b = 1$ is not enough, since the cover time of any n vertex graph by a random walk is $\Omega(n \log n)$. Suppose that at the start of each round, each particle divides in two with probability ρ. This gives an expected branching factor of $b = 1 + \rho$. In the dual BIPS process, in each round each vertex u selects with probability ρ two random neighbours (with replacement) and with probability $1 - \rho$ only one random neighbour. Vertex u will be an infected vertex in the next iteration, if it has selected an infected neighbour in the current iteration. The duality Theorem 1.3 holds for any $b = 1 + \rho$. The bounds on the cover time and the infection time given in Theorems 1.1, 1.2, 1.4 and 1.5, hold for the COBRA and BIPS processes with parameter ρ, if $0 < \rho \leq 1$ is a constant.

In the remaining part of the paper, we prove Theorems 1.4 and 1.5, assuming from now on that $b = 2$, unless stated otherwise. Our analysis of the BIPS process presented in [4] was based on estimating the expected increase of the size of A_t in one round and relying on a strong concentration of this increase during the middle stage of the process. Separate care had to be given to the initial stage of the process when A_t is still relatively small and to the final stage when $V \setminus A_t$ becomes relatively small. The notion of 'relatively small' was quantified using the eigenvalue gap $1 - \lambda$.

In Theorem 1.4 we consider any graph, requiring only that it is connected. While we can still show a positive expected increase of the size of A_t in each round in this general case, the increase is too small to give any meaningful concentration, so we need a different approach.

Our proof of Theorem 1.4 is based on 'serialising' the BIPS process. All vertices which are to decide whether they will be included in the next infected set A_{t+1} make their decisions in one parallel 'global' step (round). In our analysis, however, we view this process as if the vertices were making decisions sequentially, one vertex after the other, according to an arbitrary, but fixed, order of all vertices. This serialisation is only an artifact of analysis and does not change the BIPS process in any way. To avoid potential confusion in terminology, the term *round* will refer to one 'global' parallel step of the BIPS process, and the term *step* will refer to the action of a single vertex in the serialisation of BIPS. It turs out that at this more granular level of such small steps, the process can be modelled as a martingale sequence.

The proof of Theorem 1.5 can be viewed as based on a combination of the above approach of serialising process BIPS and the approach used in [4] to provide more detailed analysis of the initial stage of the process, when the infected sets are relatively small.

2 PRELIMINARIES

The inequality given in Lemma 2.1 is a variant of the Azuma-Hoeffding inequality for super-martingales. The proof of this lemma is a straightforward adaptation of the proof of Theorem 3 from [9]. We will need in our analysis Corollary 2.2, which we derive from Lemma 2.1.

LEMMA 2.1. *Let $Z_1, Z_2, \ldots$ be a sequence of discrete random variables such that for each $i \geq 1$, $|Z_i| \leq 1$ and*

$$\mathbf{E}(Z_i | Z_1, Z_2, \ldots, Z_{i-1}) \leq 0.$$

Then for any $\delta > 0$ and $q \geq 1$, with $S_q = \sum_{i=1}^{q} Z_i$,

$$\mathbf{P}(S_q > \delta q^{1/2}) \ < \ e^{-\delta^2/2}. \tag{3}$$

PROOF. For a discrete random variable Z such that $|Z| \leq 1$ and $\mathbf{E}(Z) \leq 0$, and for any $\alpha > 0$, $\mathbf{E}(e^{\alpha Z}) < e^{\alpha^2/2}$. Because $f(x) = e^{\alpha x}$ is a convex function, for $x \in [-1, 1]$

$$
\begin{aligned}
e^{\alpha x} & \leq \ \frac{1-x}{2} e^{-\alpha} + \frac{1+x}{2} e^{\alpha} \\
& = \ \frac{1}{2} \left(e^{\alpha} + e^{-\alpha} \right) + \frac{x}{2} \left(e^{\alpha} - e^{-\alpha} \right).
\end{aligned}
$$

Thus

$$
\begin{aligned}
\mathbf{E}(e^{\alpha Z}) & \leq \ \frac{1}{2} \left(e^{\alpha} + e^{-\alpha} \right) + \frac{\mathbf{E}(Z)}{2} \left(e^{\alpha} - e^{-\alpha} \right) \\
& \leq \ \frac{1}{2} \left(e^{\alpha} + e^{-\alpha} \right) \leq e^{\alpha^2/2}.
\end{aligned}
$$

Using this bound we have for any $q \geq 1$,

$$
\begin{aligned}
\mathbf{E}(e^{\alpha S_q}) & = \ \mathbf{E}(e^{\alpha Z_q} e^{\alpha S_{q-1}}) \\
& = \ \mathbf{E}(\mathbf{E}(e^{\alpha Z_q} e^{\alpha S_{q-1}} | Z_1, Z_2, \ldots, Z_{q-1})) \\
& = \ \mathbf{E}(e^{\alpha S_{q-1}} \mathbf{E}(e^{\alpha Z_q} | Z_1, Z_2, \ldots, Z_{q-1})) \\
& < \ e^{\alpha^2/2} \cdot \mathbf{E}(e^{\alpha S_{q-1}}) \ \leq \ e^{\alpha^2 q/2},
\end{aligned}
$$

where the last inequality follows by induction. Using Markov's inequality,

$$
\begin{aligned}
\mathbf{P}(S_q > \delta q^{1/2}) & = \ \mathbf{P} \left(e^{\alpha S_q} > e^{\alpha \delta q^{1/2}} \right) \\
& \leq \ e^{-\alpha \delta q^{1/2}} \cdot \mathbf{E} \left(e^{\alpha S_q} \right) \\
& < \ e^{-\alpha \delta q^{1/2} + \alpha^2 q/2} \ = \ e^{-\delta^2/2},
\end{aligned}
$$

where the last equality holds by setting $\alpha = \delta/q^{1/2}$. □

Lemma 2.1 says that with high probability the sum S_q of random variables does not deviate too much from its expectation. In our analysis of the BIPS process, we will need high probability that the sums S_q for all $q \geq q_0$ do not deviate too much. The following corollary derived from Lemma 2.1 will work for us.

COROLLARY 2.2. *With the same setting as in Lemma 2.1, for any $\delta > 0$, $q \geq q_0 \geq 1$ and $0 < \alpha \leq 1$, we have*

$$
\begin{aligned}
\mathbf{P}(\exists\, q \geq q_0 : \ S_q > \alpha(q - q_0) + \delta q_0^{1/2}) \\
< \ q_0 e^{-\delta^2/4} + (16/\alpha^2) e^{-\alpha^2 q_0/4}.
\end{aligned}
$$

PROOF.

$$
\begin{aligned}
\mathbf{P} & \left(\exists\, q \geq q_0 : \ S_q > \alpha(q - q_0) + \delta q_0^{1/2} \right) \\
& \leq \ \sum_{q \geq q_0} \mathbf{P} \left(S_q > \alpha(q - q_0) + \delta q_0^{1/2} \right) \\
& \leq \ \sum_{q=q_0}^{2q_0-1} \mathbf{P} \left(S_q > \delta q_0^{1/2} \right) + \sum_{q \geq 2q_0} \mathbf{P} \left(S_q > \alpha q/2 \right) \\
& \leq \ \sum_{q=q_0}^{2q_0-1} \mathbf{P} \left(S_q > (\delta/2^{1/2}) q^{1/2} \right) \\
& \quad + \sum_{q \geq 2q_0} \mathbf{P} \left(S_q > (\alpha q^{1/2}/2) q^{1/2} \right) \\
& \leq \ q_0 e^{-\delta^2/4} + \sum_{q \geq 2q_0} e^{-\alpha^2 q/8} \tag{4} \\
& \leq \ q_0 e^{-\delta^2/4} + (16/\alpha^2) e^{-\alpha^2 q_0/4}. \tag{5}
\end{aligned}
$$

Inequality (4) follows from (3) applied to each term in both sums. Inequality (5) follows from the assumption that $0 < \alpha \leq 1$, so $e^{-\alpha^2/8} \leq 1 - \alpha^2/16$. □

3 INCREMENTAL NATURE OF BIPS ON ARBITRARY GRAPHS (THEOREM 4)

We consider in this section the BIPS process on any connected graph, not necessarily regular. We will be tracking the *degree* $d(A_t)$ of the infected set. For a subset of vertices $S \subseteq V$, its degree is defined as $d(S) = \sum_{u \in S} d(u)$, where $d(u)$ is the degree of vertex u.

We consider any round $t \geq 1$ such that $d(A_{t-1}) < 2m = d(V)$, that is, any round before the completion of the process (before the infected set becomes V). Let $A = A_{t-1}$ and $B = A_t$. We view the next infected set B as the disjoint union of sets B_{fix} and B_{rand}, where B_{fix} is the 'deterministic' part of B consisting of the vertices which have all neighbours in A, and $B_{\text{rand}} = B \setminus B_{\text{fix}}$ is the 'random' part of B. Note that according to this definition, the source v is in B_{fix}, if all its neighbours are in A, and in B_{rand} otherwise. Defining $\mathcal{N}(S) = \bigcup_{u \in S} \mathcal{N}(u)$, we can write

$$
\begin{aligned}
B_{\text{fix}} & = \ \{u \in V : \ \mathcal{N}(u) \subseteq A\}, \\
B_{\text{rand}} & \subseteq \ (\mathcal{N}(A) \cup \{v\}) \setminus B_{\text{fix}} \ \equiv \ C. \tag{6}
\end{aligned}
$$

The set $C = C_t$, which contains the candidates for inclusion in set B_{rand}, is never empty. If the source v is not in B_{fix}, then $v \in C$. If $v \in B_{\text{fix}}$, that is, v and all its neighbours are in A, then in a shortest path $(v, \ldots, u', u'', x)$ from the source v to a vertex in $V \setminus A$, vertex u'' is in C.

We have

$$d(A) - E(A, C) \ = \ d(B_{\text{fix}}) \ \leq \ d(A), \tag{7}$$

where for subsets $X, Y \subseteq V$, not necessarily disjoint,

$$E(X, Y) \ \equiv \ \sum_{u \in X} d_Y(u) \ = \ \sum_{u \in Y} d_X(u),$$

and $d_Y(u) \equiv |\mathcal{N}(u) \cap Y|$ is the number of the neighbours of vertex u which belong to set Y. We have the following lower bound for

the expectation of $d(B_{\text{rand}})$.

$\mathbf{E}(d(B_{\text{rand}})) =$

$$= 1_{v \in C} \cdot d(v) + \sum_{u \in C \setminus \{v\}} d(u) \left(1 - \left(1 - \frac{d_A(u)}{d(u)} \right)^2 \right)$$

$$= 1_{v \in C} \cdot d(v) + \sum_{u \in C \setminus \{v\}} d_A(u) \left(2 - \frac{d_A(u)}{d(u)} \right)$$

$$\geq \sum_{u \in C} d_A(u) \left(2 - \frac{d_A(u)}{d(u)} \right) \geq \sum_{u \in C} d_A(u) \left(1 + \frac{1}{d(u)} \right)$$

$$\geq \left(1 + \frac{1}{d_{\max}} \right) E(A, C). \tag{8}$$

Using (7) and (8), we get the following lower bound on $\mathbf{E}(d(B))$.

$$\mathbf{E}(d(B)) \quad = \quad d(B_{\text{fix}}) + \mathbf{E}(d(B_{\text{rand}})) \tag{9}$$
$$= \quad d(A) - E(A, C) + \mathbf{E}(d(B_{\text{rand}}))$$
$$\geq \quad d(A) + \frac{1}{d_{\max}} E(A, C). \tag{10}$$

We write

$$d(B_{\text{rand}}) \quad = \quad \sum_{u \in C} d(u) X_u, \tag{11}$$

where the binary random variable X_u is equal to 1 if, and only if, $v \in B_{\text{rand}}$. If C includes the source v, then $X_v \equiv 1$. Using (7) and (11), we have

$$d(B) \quad = \quad d(B_{\text{fix}}) + d(B_{\text{rand}})$$
$$= \quad d(A) - E(A, C) + \sum_{u \in C} d(u) X_u$$
$$= \quad d(A) + \sum_{u \in C} (d(u) X_u - d_A(u)). \tag{12}$$

For $1 \leq \tau \leq t \leq T$, where T denotes the completion time of BIPS, the relation (12) for round τ can be written as

$$d(A_\tau) = d(A_{\tau-1}) + \sum_{u \in C_\tau} \left(d(u) X_{\tau, u} - d_{A_{\tau-1}}(u) \right). \tag{13}$$

Summing (13) over rounds $\tau = 1, 2, \ldots, t$ gives

$$d(A_t) = d(v) + \sum_{\tau=1}^{t} \sum_{u \in C_\tau} \left(d(u) X_{\tau, u} - d_{A_{\tau-1}}(u) \right)$$
$$\equiv \quad d(v) + \sum_{l=1}^{v(t)} Y_l. \tag{14}$$

Here $v(t) \equiv \sum_{\tau=1}^{t} |C_\tau|$ and

$$Y_{v(\tau-1)+i} \equiv d(u) X_{\tau, u} - d_{A_{\tau-1}}(u),$$

for $1 \leq i \leq |C_\tau|$, where u is the i-th smallest vertex of C_τ, according to some arbitrary but fixed ordering of the vertices in V. Since $C_\tau \neq \emptyset$, for each $1 \leq \tau \leq T$, we have

$$t \quad \leq \quad v(t) \quad \leq \quad v(t-1) + n, \tag{15}$$

setting $v(0) = 0$. We say that round $t \leq T$ consists of $|C_t|$ steps, with the random variable $Y_{v(t-1)+i}$ corresponding to step i of this round. Thus we can view the BIPS process as a sequence of steps $1, 2, \ldots, l, \ldots$, which are grouped into rounds.

While the BIPS process completes at round T, the sequence $(A_t)_{t \geq 0}$ is defined in the natural way for all $t \geq 1$: $A_t = V$ and $d(A_t) = 2m$, for each $t > T$. The sequence (Y_l) is defined for $1 \leq l \leq v(T)$, that is, up till the completion of the BIPS process. For technical convenience, we set $Y_l = 1$ for all $l > v(T)$. The choice of the value 1 will become clear later. We note that (14) holds only for $t \leq T$.

The random variables Y_l are not independent. The distribution of Y_l depends on the values of variables Y_i, $1 \leq i \leq l - 1$. For any fixed $l \geq 1$ and any sequence of numbers $y_1, y_2, \ldots, y_{l-1}$, either this sequence of numbers is not a feasible sequence of values for the sequence of variables $Y_1, Y_2, \ldots, Y_{l-1}$, or it is feasible, shows in full the evolution of the BIPS process up till step $l-1$ and defines the distribution of the variable Y_l. Indeed, if $Y_1 = y_1, Y_2 = y_2, \ldots, Y_{l-1} = y_{l-1}$, then, starting from the known initial A_0 and C_1 and knowing that the vertices of C_1 are considered according to a fixed ordering of all vertices of V, we can keep tracking the values of $Y_1, Y_2, \ldots$ to identify the vertices in A_1 (this also gives the set C_2) then the vertices in A_2, and so on. Finally, either the process has completed before step l, so $Y_l \equiv 1$, or we identify the round t which includes step l, the set $A_{t-1} \subseteq V$ of the vertices infected in the previous round, and the vertex u considered in step l. In both cases, we get the distribution of the random variable Y_l.

Equation (14) implies that instead of analysing the sequence $d(A_t)$, $t \geq 1$, we can analyse the sequence of sums $R_q = \sum_{l=1}^{q} Y_l$, $q \geq 1$. There is a technical complication here because only for those $q = v(t)$ does the value of R_q correspond to the value of $d(A_t)$. This means that a large value of some R_q does not immediately imply a large value of $d(A_t)$. However, an appropriately long sequence $R_q, R_{q+1}, \ldots, R_{q'}$ of large values would imply a large value of some $d(A_t)$. More precisely, we have the following relation between the sequences $(d(A_t))_{t \geq 1}$ and $(R_q)_{q \geq 1}$. For each $1 \leq k \leq 2m - d(v)$ and each $t \geq 1$,

$$(d(A_t) < d(v) + k) \Rightarrow \left(\exists t \leq q \leq tn : R_q < k \right). \tag{16}$$

Indeed, consider an execution of the BIPS process such that $d(A_t) < d(v) + k \leq 2m$. From (14), $R_{v(t)} = \sum_{l=1}^{v(t)} Y_l < k$, and from (15), $t \leq v(t) \leq tn$. Thus $R_q < k$, for some $t \leq q \leq tn$.

We next derive a lower bound on the conditional expectation of Y_l given the values of the variables $Y_1, Y_2, \ldots, Y_{l-1}$. If these values show that the BIPS process has already completed (that is, $l > v(T)$), then $Y_l \equiv 1$ and $\mathbf{E}(Y_l | Y_1, Y_2, \ldots, Y_{l-1}) = 1$. Otherwise, let u denote the vertex corresponding to Y_l, let t denote the index of the current round (that is, the round which includes step l) and let $A = A_{t-1}$. As mentioned above, u, t and A_{t-1} are fully determined by the values of variables $Y_1, Y_2, \ldots, Y_{l-1}$. If $u = v$ (the source vertex), then $Y_l = d(v) - d_A(v)$ and $0 \leq d_A(v) \leq d(v) - 1$, so $Y_l \geq 1$. (In this case, $v \in C$, so $v \notin B_{\text{fix}}$ and $d_A(v) \leq d(v) - 1$.) If $u \neq v$, then

$$\mathbf{E}(Y_l \mid Y_1, Y_2, \ldots, Y_{l-1})$$

$$= d(u) \left(1 - \left(1 - \frac{d_A(u)}{d(u)} \right)^2 \right) - d_A(u)$$

$$= d_A(u) \left(1 - \frac{d_A(u)}{d(u)} \right) \geq 1 - \frac{1}{d(u)}. \tag{17}$$

The inequality above holds because $u \in \mathcal{N}(A) \setminus B_{\text{fix}}$, so $1 \leq d_A(u) \leq d(u) - 1$. This also implies that $d(u) \geq 2$. Thus in all cases,

$$\mathbf{E}(Y_l | Y_1, Y_2, \ldots, Y_{l-1}) \geq \frac{1}{2}. \tag{18}$$

The proof of Theorem 1.4 follows from Lemma 3.1 (proved below) by choosing $k = 2m - d(v)$.

LEMMA 3.1. *Consider the BIPS process on a connected graph with n vertices, m edges and the maximum vertex degree $d_{\max}$. For any constant $C > 0$, there exists a constant $C' > 0$, such that for any $1 \leq k \leq 2m - d(v)$ and $t(k) = 4k + C'(d_{\max})^2 \log n$,*

$$\mathbf{P}(\exists \, t \geq t(k): \, d(A_t) < d(v) + k) \leq n^{-C}.$$

PROOF. For each $l \geq 1$, if the variable Y_l has the corresponding vertex u (that is, if $l \leq v(T)$) and u is not the source v, then $Y_l \in \{-d_A(u), d(u) - d_A(u)\}$. If u is the source, then $Y_l = d(v) - d_A(v)$. If $l > v(T)$, then $Y_l = 1$. Thus in all cases $|Y_l| \leq d_{\max} \equiv D$. We define $Z_l = (1/2 - Y_l)/D$, so that $|Z_l| \leq 1$ and, from (18), $\mathbf{E}(Z_l | Z_1, Z_2, \ldots, Z_{l-1}) \leq 0$. We will use Corollary 2.2 applied to the sum $\sum_{l=1}^{q} Z_l$.

For an arbitrary constant $C > 0$, let $\delta = \sqrt{4(C+4)\log n}$ and $C' = 16(C+4)$. These settings imply that for $q_0 = t(k) = 4k + C'D^2 \log n$, we have $q_0/2 - D\delta q_0^{1/2} \geq k$. Thus, using first the implication (16), then changing from variables Y_l to variables Z_l and using Corollary 2.2, we obtain

$$\mathbf{P}(\exists \, t \geq q_0: \, d(A_t) < d(v) + k)$$

$$\leq \quad \mathbf{P}\left(\exists \, q \geq q_0: \, \sum_{l=1}^{q} Y_l < k\right)$$

$$= \quad \mathbf{P}\left(\exists \, q \geq q_0: \, \sum_{l=1}^{q} Z_l > \frac{q}{2D} - \frac{k}{D}\right)$$

$$\leq \quad \mathbf{P}\left(\exists \, q \geq q_0: \, \sum_{l=1}^{q} Z_l > \frac{q - q_0}{2D} + \delta q_0^{1/2}\right)$$

$$\leq \quad q_0 e^{-\delta^2/4} + 64 D^2 e^{-q_0/(16D^2)} \tag{19}$$

$$\leq \quad n^3 e^{-\delta^2/4} + n^3 e^{-(C'/16)\log n} \tag{20}$$

$$= \quad n^{-(C+1)} + n^{-(C+1)} \, < \, n^{-C}. \tag{21}$$

Inequality (19) follows from the inequality stated in Corollary 2.2, taking $\alpha = 1/(2D)$. Inequality (20) follows from the definition of q_0 and Equality (21) follows from our settings for δ and C'. □

4 BIPS ON REGULAR GRAPHS: KNOWN PROPERTIES

The analysis of the BIPS process on regular graphs given in [4] was broken into three phases. The first phase brings up the infection size from 1 to $\Omega(\log n/(1-\lambda)^2)$, the second phase increases it to $\Theta(n)$, and finally the third phase deals with comleting the infection of the whole graph. The first phase is the slow one, requiring $O(\log n/(1-\lambda)^3)$ time. The second and third phases require only $O(\log n/(1-\lambda))$ time and their joint performance is summarised in Lemma 4.3. All phases use Lemma 4.1, which gives a lower bound on the expected increase of infection in one round. Lemma 4.2 is the analog of Lemma 4.1 for branching factor $b = 1 + \rho$.

To improve the $O(\log n/(1-\lambda)^3)$ bound for small $1 - \lambda$, we will restructure the analysis presented in [4], but we will still refer directly to Lemmas 4.1, 4.2 and 4.3.

LEMMA 4.1. [4] *Let G be a connected r-regular graph on n vertices, with $\lambda < 1$, where λ is the absolute second eigenvalue of the random-walk transition matrix. Let A_t be the size of the infected set after round t of the BIPS process with $b = 2$, then*

$$\mathbf{E}(|A_{t+1}| \mid A_t = A) \geq |A|(1 + (1 - \lambda^2)(1 - |A|/n)).$$

LEMMA 4.2. [4] *Let A_t be the size of the infected set after round t of the BIPS process with expected branching factor $b = 1 + \rho$, then*

$$\mathbf{E}(|A_{t+1}| \mid A_t = A) \geq |A|(1 + \rho(1 - \lambda^2)(1 - |A|/n)).$$

LEMMA 4.3. [4] *Let G be a connected n-vertex r-regular graph and consider the BIPS process on G from some round $t > 0$. There exist constants C and K such that, if $1 - \lambda \geq C\sqrt{\log n/n}$ and $|A_t| \geq K \log n/(1-\lambda)^2$, then the whole graph is infected within $O(\log n/(1-\lambda))$ additional rounds with probability at least $1 - O(1/n^3)$.*

5 BIPS ON REGULAR GRAPHS: NEW ANALYSIS

In comparison with the analysis given in [4] and outlined in the previous section, our new analysis ends the first, initial phase earlier, more precisely as soon as the size of infection becomes (roughly) $\Omega(\log n/(1-\lambda))$. From that point we analyse in more detail how the size of infection increases until it reaches $\Theta(n)$. In this analysis we use and extend the methodology developed in Section 3. The analysis of the final phase, from infection of size $\Theta(n)$ to complete infection of the whole graph, is the same as in [4], so we simply refer to Lemma 4.3.

When considering regular graphs, for simplicity we track the size of the current infection set rather than the degree of this set. Applying Lemma 3.1 to r-regular graphs and substituting k with $(\kappa - 1)r$, we obtain the following corollary.

COROLLARY 5.1. *Consider the BIPS process on a connected r-regular graph with n vertices. For any constant $C > 0$, there exists a constant $C' > 0$, such that for any $1 \leq \kappa \leq n$ and $t(\kappa) = 4r\kappa + C'r^2 \log n$,*

$$\mathbf{P}(\exists \, t \geq t(\kappa): \, |A_t| < \kappa) \leq n^{-C}. \tag{22}$$

Consider r-regular graphs with the eigenvalue gap $1 - \lambda \geq C\sqrt{\log n/n}$, where C is the constant from Lemma 4.3. By applying Corollary 5.1 with $\kappa = K \log n/(1-\lambda)^2$ and then Lemma 4.3, we conclude that BIPS completes within $O((r/(1-\lambda)^2 + r^2) \log n)$ rounds, w.h.p. and in expectation. The constant K here is as needed in Lemma 4.3. In the remaining part of this section we extend our analysis to reduce this bound to $O((r/(1-\lambda) + r^2) \log n)$. To achieve this, we need to bridge the gap between the infection size $\Omega((\log n)/(1-\lambda))$ guaranteed by Corollary 5.1 after $O((r/(1-\lambda) + r^2) \log n)$ rounds, and the initial infection size $\Omega(\log n/(1-\lambda)^2)$ required by Lemma 4.3.

Corollary 5.1 with $\kappa = \Omega(r \log n)$ gives the size of infection $\Omega(\kappa)$ within $t(\kappa) = O(r\kappa)$ rounds. Thus the infection grows with the average rate of $\Omega(1/r)$ vertices per round. In the general case, that is, for any structure of regular graphs, we can show only this small rate of growth, because we can only guarantee that each set C_t (see (6)) contains at least one vertex. However, when the size of

infection keeps increasing and passes through some appropriate thresholds, then, depending on the structure of the graph, the sizes of sets C_t also grow, gradually speeding up the rate of growth of infection. While there are various ways of lower bounding the size of C_t in terms of the size of A_t and the eigenvalue gap $1 - \lambda$ of the graph, the following bound follows easily from the facts which we have already established.

COROLLARY 5.2. *For each round $t \geq 1$ of the BIPS process on an n-vertex r-regular graph, if $|A_{t-1}| \leq n/2$, then the size of the set C_t defined in (6) is at least $|A_{t-1}|(1 - \lambda)/2$.*

PROOF. We use the same notation as in Section 3: $A \equiv A_{t-1}$, $B \equiv A_t$ and $C \equiv C_t$. We have

$$|A| + |A|(1 - \lambda)/2 \;\; \leq \;\; \mathbf{E}(|B|) \;\; = \;\; |B_{\text{fix}}| + \mathbf{E}(|B_{\text{rand}}|)$$
$$\leq \;\; |A| + \mathbf{E}(|B_{\text{rand}}|).$$

The first inequality follows from Lemma 4.1 and the assumption that $|A| \leq n/2$. The middle equality is (9) divided by r, and the last inequality is (7) divided by r. Thus $\mathbf{E}(|B_{\text{rand}}|) \geq |A|(1 - \lambda)/2$, while $|C|$ is an obvious upper bound on $\mathbf{E}(|B_{\text{rand}}|)$ because $B_{\text{rand}} \subseteq C$.   □

We consider now the BIPS process from some round $t \geq 1$ and denote by $\tilde{Y}_l$ and $\tilde{R}_q$ the random variables Y_l and R_q from Section 3 divided by r. In this setting, (14) and (16) become

$$|A_{t+\Delta}| = |A_t| + \sum_{l=1}^{\nu(\Delta)} \tilde{Y}_l \;\; \equiv \;\; |A_t| + \tilde{R}_{\nu(\Delta)}, \qquad (23)$$

$$(|A_{t+\Delta}| < |A_t| + \kappa) \Rightarrow (\exists\, \Delta \leq q < \Delta n : \tilde{R}_q < \kappa), \qquad (24)$$

where $1 \leq \kappa \leq n - |A_t|$ and $\Delta \geq 1$. Generalising (24), we obtain the following implication, where $\alpha \geq 1$ is the threshold on the sizes of sets $C_{t+\tau}$ and $\mathcal{A}$ is an arbitrary event.

$$((|A_{t+\Delta}| < |A_t| + \kappa) \wedge \mathcal{A})$$
$$\Rightarrow \left(\left(\exists\, \alpha\Delta \leq q < \Delta n : \tilde{R}_q < \kappa \right) \right)$$
$$\text{or } ((\exists\, 1 \leq \tau \leq \Delta : |C_{t+\tau}| < \alpha) \wedge \mathcal{A}). \qquad (25)$$

Indeed, if $|A_{t+\Delta}| < |A_t| + \kappa$ and $|C_{t+\tau}| \geq \alpha$ for each $1 \leq \tau \leq \Delta$ (that is, each round $\tau = t+1, t+2, \ldots, t+\Delta$ has at least α steps), then $\tilde{R}_{\nu(\Delta)} < \kappa$ and $\alpha\Delta \leq \nu(\Delta) < \Delta n$. Following the same argument as in the proof of Lemma 3.1 but for regular graphs and using (25) instead of (24), Corollary 5.1 becomes the following statement.

COROLLARY 5.3. *Consider the BIPS process on a connected r-regular graph with n vertices. For any constant $C > 0$, there exists a constant $C' > 0$, such that for any $\alpha \geq 1$, $t \geq 1$, infected set A_t ($v \in A_t \subset V$), $1 \leq \kappa \leq n - |A_t|$ and an event $\mathcal{A}$, defining*

$$\Delta(\kappa, \alpha) = (4r\kappa + C'r^2 \log n)/\alpha, \qquad (26)$$

we have

$$\mathbf{P}((\exists\, \Delta \geq \Delta(\kappa, \alpha) : |A_{t+\Delta}| < |A_t| + \kappa) \wedge \mathcal{A})$$
$$\leq n^{-C} + \mathbf{P}((\exists\, 1 \leq \tau \leq \Delta(\kappa, \alpha) : |C_{t+\tau}| < \alpha) \wedge \mathcal{A}).$$

Using Corollaries 5.1, 5.2 and 5.3, we obtain the following lemma.

LEMMA 5.4. *Consider the BIPS process on a connected n-vertex r-regular graph G with eigenvalue gap $1 - \lambda$. For any constant $C > 0$, there exists a constant C'' such that the probability that the size of*

infection is at least $n/4$ at some round $t \leq C''r(1/(1 - \lambda) + r) \log n$ is at least $1 - n^{-C}$.

PROOF. Let $C > 0$ be an arbitrary constant and let $C' > 0$ be a constant which 'works' for the constant $C + 2$ in both Corollaries 5.1 and 5.3. We first apply Corollary 5.1 with $k = \kappa_0$, where

$$\kappa_0 = \min\{1/(1 - \lambda) + (C'r/4) \log n, \; n\},$$

and conclude that with probability at least $1 - n^{-(C+2)}$, the size of infection is at least κ_0 in each round $t \geq t_0 = 8r\kappa_0$. Reaching infection size κ_0 can be viewed as the first, initial phase of the BIPS process. If $\kappa_0 \geq n/4$, then we are done. Otherwise, we will be repeatedly doubling the target size of infection, that is, we will be analysing when the infection size is w.h.p. at least $\kappa_i = 2^i \kappa_0$, for $i = 1, 2, \ldots, j$, where $n/4 \leq 2^j \kappa_0 < n/2$.

Let $t_i = t_0 + 16ir/(1 - \lambda)$ and let $\mathcal{A}$ be the event that $|A_t| < n/2$ for all $1 \leq t \leq t_j$. Observe that $t_{i+1} \geq t_i + \Delta(\kappa_i, \alpha_i)$, where $\Delta(\kappa, \alpha)$ is as defined in (26) and $\alpha_i = \kappa_i(1 - \lambda)/2$. We prove by induction that for each $0 \leq i \leq j$,

$$\mathbf{P}((\exists\, t \geq t_i : |A_t| < \kappa_i) \wedge \mathcal{A}) < 3^i n^{-(C+2)}. \qquad (27)$$

Observe that (27) with $i = j$ implies that the probability that $|A_t| < n/4 \leq \kappa_j$, for each $1 \leq t \leq t_j = O(r(1/(1 - \lambda) + r) \log n)$, is at most $3^j n^{-(C+2)} < n^{-C}$, as claimed in the lemma. Thus it only remains to prove (27).

For $i = 0$, Inequality (27) is established in the first paragraph of the proof. Let $0 \leq i \leq j - 1$ and consider (27) for $i + 1$. We have, using the equivalent notation $A(t) \equiv A_t$,

$$\mathbf{P}((\exists\, t \geq t_{i+1} : |A_t| < \kappa_{i+1}) \wedge \mathcal{A})$$
$$\leq \mathbf{P}((\exists\, \Delta \geq \Delta(\kappa_i, \alpha_i) : |A(t_i + \Delta)| < \kappa_{i+1}) \wedge \mathcal{A})$$
$$\leq \max\{\, \mathbf{P}((\exists\, \Delta \geq \Delta(\kappa_i, \alpha_i) : |A(t_i + \Delta)| < \kappa_{i+1}) \wedge \mathcal{A} \mid A_{t_i}) :$$
$$v \in A(t_i) \subseteq V, |A(t_i)| \geq \kappa_i\}$$
$$+ \mathbf{P}((|A(t_i)| < \kappa_i) \wedge \mathcal{A}). \qquad (28)$$

The induction hypothesis implies that

$$\mathbf{P}((|A(t_i)| < \kappa_i) \wedge \mathcal{A}) \leq 3^i n^{-(C+2)}. \qquad (29)$$

If at the current round t we reduce the infection set A_t to $A'_t \subseteq A_t$, then for any $t' \geq t$, the distribution of $|A'_{t'}|$ stochastically dominates the distribution of $|A_{t'}|$. Thus, if $v \in A' \subseteq A'' \subseteq V$ and $|A'| = \kappa_i$, then

$$\mathbf{P}((\exists\, \Delta \geq \Delta(\kappa_i, \alpha_i) : |A(t_i + \Delta)| < \kappa_{i+1}) \wedge \mathcal{A} \mid A(t_i) = A'')$$
$$\leq \mathbf{P}((\exists\, \Delta \geq \Delta(\kappa_i, \alpha_i) : |A(t_i + \Delta)| < \kappa_{i+1}) \wedge \mathcal{A} \mid A(t_i) = A'),$$

so it is enough to consider in 'max' in (28) only sets $A(t_i)$ of size κ_i. For any fixed $A(t_i)$ such that $v \in A(t_i) \subseteq V$ and $|A(t_i)| = \kappa_i$,

$$\mathbf{P}((\exists\, \Delta \geq \Delta(\kappa_i, \alpha_i) : |A(t_i + \Delta)| < \kappa_{i+1}) \wedge \mathcal{A})$$
$$= \mathbf{P}((\exists\, \Delta \geq \Delta(\kappa_i, \alpha_i) : |A(t_i + \Delta)| < |A_{t_i}| + \kappa_i) \wedge \mathcal{A})$$
$$\leq \mathbf{P}((\exists\, 1 \leq \tau \leq \Delta(\kappa_i, \alpha_i) : |C(t_i + \tau)| < \alpha_i) \wedge \mathcal{A}),$$
$$+ n^{-(C+2)}. \qquad (30)$$

The inequality above follows from Corollary 5.3.

$$\mathbf{P}\left((\exists\, 1 \le \tau \le \Delta(\kappa_i, \alpha_i) : \ |C(t_i + \tau)| < \alpha_i) \wedge \mathcal{A}\right)$$
$$\le \quad \mathbf{P}\left((\exists\, t_i \le t \le t_{i+1} : |C_t| < \alpha_i) \wedge \mathcal{A}\right)$$
$$\le \quad \mathbf{P}\left((\exists\, t_i \le t \le t_{i+1} : (|A_t| < \kappa_i) \vee (|A_t| > n/2)) \wedge \mathcal{A}\right)$$
$$= \quad \mathbf{P}\left((\exists\, t_i \le t \le t_{i+1} : (|A_t| < \kappa_i)) \wedge \mathcal{A}\right)$$
$$\le \quad 3^i n^{-(C+2)} \tag{31}$$

The second inequality above follows from Corollary 5.2 and the last inequality follows from the induction hypothesis. Summarising, if we start from (28) and use (29)–(31), we obtain

$$\mathbf{P}((\exists\, t \ge t_{i+1} : |A_t| < \kappa_{i+1}) \wedge \mathcal{A})$$
$$\le 2 \cdot 3^i n^{-(C+2)} + n^{-(C+2)}$$
$$\le 3^{i+1} n^{-(C+2)}.$$

Thus (27) holds for $i+1$, so by induction, it holds for all $0 \le i \le j$. $\quad\square$

Proof of Theorem 1.5. Apply Lemma 5.4 to show that the size of infection is at least $n/4$ within $O(r(1/(1-\lambda)+r)\log n)$ rounds. Then apply Lemma 4.3 to show that the whole graph is infected after additional $O((\log n)/(1-\lambda))$ rounds. $\quad\square$

6 BIPS WITH BRANCHING FACTOR LESS THAN 2

The analysis of the BIPS process with branching factor $b = 1 + \rho$, for a constant $0 < \rho \le 1$, requires somewhat tedious but otherwise straightforward tracing of the analysis of the main case $b = 2$ and updating the probability that a vertex u catches infection in the current round. Denoting as before $A \equiv A_{t-1}$ and $B \equiv B_t$, if the branching factor is $b = 2$, then the probability that a vertex u gets infected in round t is equal to

$$\mathbf{P}_{(2)}(u \in B) = \left(1 - \left(1 - \frac{d_A(u)}{d(u)}\right)^2\right). \tag{32}$$

If the branching factor is $b = 1 + \rho \le 2$, then this probability is

$$\mathbf{P}_{(1+\rho)}(u \in B) = \left(1 - \left(1 - \frac{d_A(u)}{d(u)}\right)\left(1 - \rho\frac{d_A(u)}{d(u)}\right)\right). \tag{33}$$

We give a couple of examples of substituting (32) with (33) in the analysis. In Section 3, the probabilities of vertices becoming infected are used in (8) – (10) to obtain a lower bound on the expected size of infection in the next round. Adapting to the case $b = 1 + \rho \le 1$ gives:

$$\mathbf{E}(d(B_{\text{rand}})) =$$
$$= 1_{v \in C} \cdot d(v) + \sum_{u \in C \setminus \{v\}} d(u)\left(1 - \left(1 - \frac{d_A(u)}{d(u)}\right)\left(1 - \rho\frac{d_A(u)}{d(u)}\right)\right)$$
$$= 1_{v \in C} \cdot d(v) + \sum_{u \in C \setminus \{v\}} d_A(u)\left(1 + \rho - \rho\frac{d_A(u)}{d(u)}\right)$$
$$\ge \sum_{u \in C} d_A(u)\left(1 + \rho - \rho\frac{d_A(u)}{d(u)}\right) \ge \left(1 + \frac{\rho}{d_{\max}}\right)E(A,C),$$

and

$$\mathbf{E}(d(B)) \quad \ge \quad d(A) + \frac{\rho}{d_{\max}}E(A,C).$$

Similarly (17) and (18) become:

$$\mathbf{E}(Y_l \mid Y_1, Y_2, \ldots, Y_{l-1}) \ge \ \rho\left(1 - \frac{1}{d(u)}\right),$$

and

$$\mathbf{E}(Y_l | Y_1, Y_2, \ldots, Y_{l-1}) \ \ge \ \frac{\rho}{2}.$$

The updates are also needed in Lemma 3.1 and Corollaries 5.1 and 5.3, where the number of rounds guaranteeing w.h.p. the increase of the size of infection (numbers $t(k)$, $t(\kappa)$ and $\Delta(\kappa, \alpha)$) should be multiplied by $1/\rho^2$.

7 CONCLUSIONS

The COBRA process was proposed for studies as a type of parallel random walks with relations to epidemics processes and potential applications in network algorithms [6, 8]. We have contributed to this studies by improving upper bounds on the cover time of the COBRA process. In particular we achieved a significant improvement of the general bound which applies to all connected graphs. We achieved this by exploiting the duality between COBRA and a related epidemics process.

The obvious open questions are about tightness of the existing bounds. For example, while our general bound of $O(n^2 \log n)$ is a significant improvement over the previous best bound of $O(n^{11/4} \log n)$, there are no known examples of the cover time $\omega(n \log n)$. It has actually been conjectured the worst-case cover time for any graph is $O(n \log n)$. Regarding cover time bounds for specific classes of graphs, the hypercube remains an interesting example, with the best known upper bound of $O(\log^3 n)$, implied by the results in this paper, and no good reason why it should be higher than $\Theta(\log n)$.

REFERENCES

[1] N. Alon, C. Avin, M. Koucký, G. Kozma, Z. Lotker, and M. R. Tuttle. Many random walks are faster than one. *Combinatorics, Probability & Computing*, 20(4):481–502, 2011.

[2] A. Z. Broder, A. R. Karlin, P. Raghavan, and E. Upfal. Trading space for time in undirected $s-t$ connectivity. In *STOC*, pages 543–549. ACM, 1989.

[3] C. Cooper, A. M. Frieze, and T. Radzik. Multiple random walks in random regular graphs. *SIAM J. Discrete Math.*, 23(4):1738–1761, 2009.

[4] C. Cooper, T. Radzik, N. Rivera. The Coalescing-Branching Random Walk on Expanders and the Dual Epidemic Process. PODC 2016: 461-467.

[5] C. Dutta, G. Pandurangan, R. Rajaraman, and S. Roche. Coalescing-branching random walks on graphs. In *Proceedings of the Twenty-fifth Annual ACM Symposium on Parallelism in Algorithms and Architectures*, SPAA '13, pages 176–185, New York, NY, USA, 2013. ACM.

[6] C. Dutta, G. Pandurangan, R. Rajaraman, and S. Roche. Coalescing-branching random walks on graphs. *ACM Trans. Parallel Comput.*, 2(3):20:1–20:29, Nov. 2015.

[7] R. Elsässer and T. Sauerwald. Tight bounds for the cover time of multiple random walks. In *Automata, Languages and Programming, 36th International Colloquium, ICALP 2009, Proceedings, Part I*, volume 5555 of *Lecture Notes in Computer Science*, pages 415–426. Springer, 2009.

[8] M. Mitzenmacher, R. Rajaraman, S. Roche. Better bounds for coalescing-branching random walks. In *Proceedings of the Twenty-eight Annual ACM Symposium on Parallelism in Algorithms and Architectures*, SPAA '16, pages 313–323, 2016. (see also http://arxiv.org/abs/1603.06109)

[9] E. Shamir and J. Spencer. Sharp concentration of the chromatic number on random graphs $G_{n,p}$. *Combinatorica*, March 1987, Volume 7, Issue 1, pp 121-129.