View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Apollo

Spread of Information and Diseases via Random Walks in Sparse Graphs

George Giakkoupis
Inria, Univ Rennes, CNRS, IRISA, Rennes, France
george.giakkoupis@inria.fr

Hayk Saribekyan
University of Cambridge, UK
hs586@cam.ac.uk

Thomas Sauerwald
University of Cambridge, UK
thomas.sauerwald@cl.cam.ac.uk

Abstract

We consider a natural network diffusion process, modeling the spread of information or infectious diseases. Multiple mobile agents perform independent simple random walks on an n-vertex connected graph G. The number of agents is linear in n and the walks start from the stationary distribution. Initially, a single vertex has a piece of information (or a virus). An agent becomes informed (or infected) the first time it visits some vertex with the information (or virus); thereafter, the agent informs (infects) all vertices it visits. Giakkoupis et al. [16] have shown that the spreading time, i.e., the time before all vertices are informed, is asymptotically and w.h.p. the same as in the well-studied randomized rumor spreading process, on any d-regular graph with $d = \Omega(\log n)$. The case of sub-logarithmic degree was left open, and is the main focus of this paper.

First, we observe that the equivalence shown in [16] does not hold for small d: We give an example of a 3-regular graph with logarithmic diameter for which the expected spreading time is $\Omega(\log^2 n/\log\log n)$, whereas randomized rumor spreading is completed in time $\Theta(\log n)$, w.h.p. Next, we show a general upper bound of $\tilde{O}(d \cdot \mathrm{diam}(G) + \log^3 n/d)$,[1] w.h.p., for the spreading time on any d-regular graph. We also provide a version of the bound based on the average degree, for non-regular graphs. Next, we give tight analyses for specific graph families. We show that the spreading time is $O(\log n)$, w.h.p., for constant-degree regular expanders. For the binary tree, we show an upper bound of $O(\log n \cdot \log\log n)$, w.h.p., and prove that this is tight, by giving a matching lower bound for the cover time of the tree by n random walks. Finally, we show a bound of $O(\mathrm{diam}(G))$, w.h.p., for k-dimensional grids ($k \geq 1$ is constant), by adapting a technique by Kesten and Sidoravicius [22, 23].

2012 ACM Subject Classification Mathematics of computing → Stochastic processes; Theory of computation → Random walks and Markov chains; Mathematics of computing → Graph algorithms

Keywords and phrases parallel random walks, information dissemination, infectious diseases

Digital Object Identifier 10.4230/LIPIcs.DISC.2020.4

Related Version A full version of the paper is available at `https://hal.inria.fr/hal-02913942`.

Funding *George Giakkoupis*: Supported in part by ANR Project PAMELA (ANR16-CE23-0016-01).
Hayk Saribekyan: Gates Cambridge Scholarship programme.
Thomas Sauerwald: Supported by the ERC Grant 'Dynamic March'.

[1] The tilde notation hides factors of order at most $(\log\log n)^2$.

© George Giakkoupis, Hayk Saribekyan, and Thomas Sauerwald;
licensed under Creative Commons License CC-BY
34rd International Symposium on Distributed Computing (DISC 2020).
Editor: Hagit Attiya; Article No. 4; pp. 4:1–4:17
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

We consider the following natural diffusion process on a connected n-vertex graph G. A collection of mobile agents perform independent parallel (discrete-time) random walks on G, starting from the stationary distribution. Initially, there is a piece of information at some arbitrary source vertex. An agent learns the information the first time it visits some informed vertex (the vertex may have received the information in the same or a previous round). From that point on, the agent spreads the information to all vertices it visits. We study the time it takes before all vertices have been informed. We will refer to this process as VISIT-EXCHANGE, following the terminology of [16].

The above process suggests a simple message broadcasting algorithm for networks: Vertices correspond to processes, and agents are tokens circulated in the network. In each round, every process sends each of the tokens it received in the previous round to a random neighbor, and if the process knows the message, it transmits the message along with each token. As observed in [16], when the number of agents/tokens is linear in n, this algorithm has similar per round message complexity as standard randomized rumor spreading [14, 21], but in several graphs it outperforms the latter, due to a more fair bandwidth utilization: each edge is equally likely to be used in each round.

A second potential application of VISIT-EXCHANGE is as a basic model for the spread of diseases in populations. One can think of agents as the members of the population, where an infected member can transmit the infection to another either by direct contact, or indirectly. In the latter case a healthy individual contracts the virus by being in a place previously visited by an infected individual [25]. Alternatively, one can think of a larger population, residing on the vertices of the graphs (e.g., vertices are cities), and a few mobile individuals are responsible for transmitting the infection between different cities. Our basic model assumes perfect contagion and no recovery. It is an interesting future direction to analyze a refined model that allows probabilistic transmission and recovery.

Several works have studied the spread of information (or viruses) via mobile agents, performing random walks or more general jump processes, in discrete or continuous time, on various families of graphs [3, 8, 10, 16, 19, 20, 23, 24, 26, 28] (see Sect. 2 for an overview of this literature). In almost all of these works, the information is transmitted only between agents when they meet at a vertex, and vertices do not store information.

The work closest to ours is [16] (see also [17]). The authors consider VISIT-EXCHANGE with $\Theta(n)$ agents, starting from stationarity, and compare the spreading time to that of randomized rumor spreading [14, 21]. In the latter protocol, information is transmitted between adjacent vertices, without the use of agents, by having each vertex communicate with a random neighbor in each round. It was observed in [16] that there are graphs in which VISIT-EXCHANGE is significantly faster than randomized rumor spreading (logarithmic versus linear spreading time), and examples where the converse is true.

A main result of [16] is that on any d-regular graph with sufficiently large degree $d = \Omega(\log n)$, VISIT-EXCHANGE and randomized rumor spreading have the same asymptotic spreading time. The intuition for this result is the following. We have that: (i) the average number of agents per vertex is constant, since there are $\Theta(n)$ agents in total, (ii) all agents start from stationarity, and (iii) the graph is regular. It follows that, in every round, a constant expected number of agents depart from each vertex, to random neighbors. This should have a similar effect in the spread of information as randomized rumor spreading, where each vertex communicates with a random neighbor in each round.

The intuition above is not hard to formalize, and prove that VISIT-EXCHANGE is at least as fast as rumor spreading asymptotically:[2] If $d \geq c \log n$, for a large enough constant c, a Chernoff bound together with a union bound show that, w.h.p., for every vertex u and round $t \leq \mathrm{poly}(n)$, at least $\Omega(d)$ agents visit the neighbors of u in round $t-1$. Thus, at least one agent visits u in round t, with constant probability. This argument, however, does not extend to the case of $d = O(\log n)$. It was thus left as an open problem in [16], whether the same result holds for graphs of degree $d = O(\log n)$.

Our Contribution. First, we answer the above open question from [16] in the negative.

▶ **Observation 1.** *There is a* 3*-regular graph G with n vertices and diameter* $\Theta(\log n)$*, such that the expected spreading time of* VISIT-EXCHANGE *on G is* $\Omega(\log^2 n / \log\log n)$*.*

The spreading time of randomized rumor spreading is $\Theta(\mathrm{diam}(G))$, w.h.p., on any constant degree graph G [14], thus it is logarithmic for the graph above. To simplify the exposition, here we only give an example of a constant-degree, *non-regular* graph G with diameter and spreading time as described in Observation 1. Consider a 3-regular graph R with n vertices and diameter $\Theta(\log n)$ (e.g., a 3-regular expander), and $\sqrt{n}$ path graphs, each of length $\log n/2$. We obtain G by connecting one of the two endpoints of each path graph, to a distinct vertex of R. The diameter of G is clearly logarithmic. The expected spreading time is $\Omega(\log^2 n / \log\log n)$, because with constant probability, at least one path graph P contains no agents initially, and then it takes $\Omega(\log^2 n / \log\log n)$ rounds before the endpoint of P not connected to R gets informed. We can replace the paths of length $\log n/2$ with "ladder" graphs, as detailed in [18], to construct a *regular* graph satisfying Observation 1.

A consequence of Observation 1 is that known bounds for rumor spreading do not readily apply to VISIT-EXCHANGE for low-degree regular graphs, thus new bounds are needed. In view of that, we first provide a general upper bound for VISIT-EXCHANGE for regular graphs of degree $d = O(\log n)$, in terms of the graph diameter. Then we provide tight bounds for several interesting graph families. All our results assume that the number of agents is αn, for some arbitrary constant $\alpha > 0$, and the walks start from the stationary distribution. We denote by $T(G)$ the spreading time on graph G. Since all our bounds hold for any source vertex, we do not explicitly specify a source in the notation. Moreover, we omit G and write just T, when the graph is clear from the context. We write *w.h.p. (with high probability)* to denote a probability that is at least $1 - n^{-c}$ for some constant $c > 0$.

▶ **Theorem 2.** *For any d-regular graph G with* $d = O(\log n)$*,* $T = \tilde{O}(d \cdot \mathrm{diam}(G) + \log^3 n/d)$*, w.h.p., where the tilde notation hides factors of order at most* $(\log\log n)^2$*.*

In the above bound, the dependence on the diameter is best possible (e.g., the spreading time along a cycle of d-cliques is proportional to the path length multiplied by d). An additive term is also needed when the diameter is sub-logarithmic, but it is not clear whether the term $\log^3 n/d$ is tight. Recall that the corresponding upper bound for randomized rumor spreading shown in [14] is $O(d \cdot (\mathrm{diam}(G) + \log n))$. Thus, it would be reasonable to guess that the right additive term is $d \cdot \log n$. However, the example in Observation 1 shows that the term must be at least $\tilde{\Omega}(\log^2 n)$. We conjecture that the tight bound is $\tilde{O}(d \cdot \mathrm{diam}(G) + \log^2 n)$.

The proof of Theorem 2 bounds the time that the information takes to spread along a given (shortest) path in the graph. We divide time into phases of length $\log^2 n$ rounds, and in each phase, we lower-bound the probability that the information spreads along a sub-path of

[2] The proof of the other direction, that rumor spreading is at least as fast as VISIT-EXCHANGE, is significantly more involved.

length $\tilde{\Omega}(\log^2 n/d)$. For $d = \omega(\log\log n)$, we show this probability to be $1-e^{-\Omega(d)}$. Moreover, we ensure that this probability bound holds, essentially, *independently* of previous phases, by considering every other phase. We prove the bound by showing a concentration result on the number of agents at the neighborhood of each individual vertex in the sub-path, at each round of the phase, and then applying a union bound. To boost the above probability to $1 - e^{-\Omega(\log n)}$, we need $\log n/d$ phases, which yields the $\log^3 n/d$ term of the bound. For the case of $d = O(\log\log n)$, we use a similar approach, but argue instead about the number of agents that visit each vertex in the sub-path over an interval of multiple rounds (instead of looking at its neighborhood at each round). The main technical tool we use is an upper bound on the return probability from [27].

For non-regular graphs, a similar analysis as for Theorem 2 yields the following result.

▶ **Theorem 3.** *For any graph G with average degree d_{avg} and minimum degree $d_{\min} = \Omega(d_{\text{avg}})$, $T = O(d_{\text{avg}} \cdot \log^2 n \cdot (\text{diam}(G) + \log n))$, w.h.p.*

Even though this bound is likely not tight, it is interesting because there is no analogue of it for randomized rumor spreading. For example, in the graph consisting of two stars with their centers connected by an edge [16], for which $d_{\text{avg}} = O(1)$, randomized rumor spreading takes linear expected time, whereas VISIT-EXCHANGE takes logarithmic time w.h.p. (and Theorem 3 gives a $\text{poly}(\log n)$ bound).

Next we show that the spreading time on expanders is optimal, i.e., logarithmic.

▶ **Theorem 4.** *For any d-regular expander G with $d \geq 3$ constant, $T = O(\log n)$, w.h.p.*

Unlike the proof of Theorem 2, where we argue about individual vertices, to prove Theorem 4 we argue about the set of all informed vertices at time t, precisely, the subset S_t of informed vertices with at least one uninformed neighbor. By the expansion property, S_t contains at least a constant fraction of all informed vertices. We claim that a constant fraction of vertices in S_t are visited by some agent between rounds t and $t + r$, w.h.p., for any t and a large enough constant r. Since d is constant, this implies that the number of informed vertices increases by a constant factor every r rounds. To prove the above claim, we argue that the probability a given agent visits S_t between t and $t + r$ is proportional to $k = |S_t|$ and r. Thus, S_t is not visited by sufficiently many agents in these r rounds with probability decreasing exponentially in $r \cdot k$. Next, we consider all possible instantiations of S_t, and apply a union bound. Since the set of informed vertices at any time is connected, the number of different instantiations of S_t can be bounded by $d^{\Theta(k)}$. Since d is constant, the claim follows by choosing constant r large enough.

We currently do not know how to extend Theorem 4 to regular expanders of degree $\omega(1) \leq d \leq O(\log n)$ (for $d = \Omega(\log n)$, the result follows from [16]).

Next we study trees. Let $R_{b,h}$ denote a rooted b-ary tree where each vertex at distance less than h from the root has b children and all leaves are at distance h from the root. The total number of vertices is $n = (b^{h+1} - 1)/(b - 1)$.

▶ **Theorem 5.** *For any b-ary tree $R_{b,h}$ with $b \geq 2$, $T = O(h\log h + \log n)$, w.h.p. Furthermore, for the binary tree $R_{2,h}$, $T = \Omega(h \log h) = \Omega(\log n \cdot \log\log n)$, w.h.p.*

Note that the spreading time on $R_{b,h}$ of the push-only version of randomized rumor spreading is $\Theta(b \log n)$, w.h.p. Thus, VISIT-EXCHANGE is slower than push for small b, and faster for larger b. Another interesting implication of Theorem 5 is that the cover time of the tree by n random walks starting from stationarity has a super-linear speedup, compared to the cover time for a single random walk, which is $\Omega(n \log^2 n)$. Our analysis suggests a deeper connection between the cover time (or other quantities) of multiple random walks and the spreading time of VISIT-EXCHANGE, which might deserve further study.

We give now an overview of the proof of Theorem 5, for the binary tree case; the case of $b > 2$ is similar. To prove the upper bound, we fix a path between the root r and a vertex u at distance at most $h - \log h$. We show that information spreads between r and u in at most $O(\log n)$ rounds w.h.p., by showing that agents arrive at each vertex v of the path at roughly constant rate, independently of the other vertices in the path. To achieve this independence, for each v we identify a subset S_v of the descendants of v at distance $\log h$, and count only visits to v by agents that are in S_v a number of $\Theta(\log h)$ rounds ago, and in the meantime have not walked past v. To show a constant rate, instead of $1/\Theta(\log h)$, a careful pipeline argument is used. To bound the time to spread the information in the last $\log h$ levels of the tree, we bound the cover time of a tree of height $\log h$ by h walks starting from the root, which takes $O(h \log h)$ steps w.h.p. (in n). Finally, to show the lower bound of Theorem 5, we bound from below the cover time of the tree by n random walks starting from stationarity.

Last we show that the spreading time on grids is optimal, i.e., asymptotically equal to the diameter. Let $G_{k,n}$ denote the k-dimensional grid with side length $n^{1/k}$ and n vertices in total (for simplicity, we assume $n^{1/k}$ is an integer).

▶ **Theorem 6.** *For any grid graph $G_{k,n}$ that has a constant number of dimensions $k \geq 1$, $T = \Theta(\mathrm{diam}(G))$, w.h.p.*

A weaker version of this result, with additional $\log \log n$ factors, follows from Theorem 2. To get rid of these extra factors, we employ a much more fine-grained analysis.

Our proof of Theorem 6 uses a technique developed by Kesten and Sidoravicius [22, 23], who proved a similar bound for a continuous-time diffusion process, in which information spreads between agents when they meet (it is not stored on vertices). For our discussion here, we assume the 1-dimensional case, i.e., the n-path. We consider a sequence of $\Theta(\log \log n)$ tessellations of space-time (up to time linear in n), where each tessellation consists of square blocks; the length of the block side is constant in the first tessellation, and increases exponentially in each subsequent tessellation. Let Δ be the side length of a block, and let (v, t) be its bottom left corner; i.e., the block contains all points $(v + j, t + j')$, $0 \leq j, j' < \Delta$. Roughly speaking, the block is "good" if a sufficiently large neighborhood of the vertices in the block (namely, vertices $v - 3\Delta$ up to $v + 4\Delta$) is sufficiently densely populated by agents at time $t - \Delta$. This implies that any space-time point in the block has a good probability of containing some agent. Starting from the last tessellation, for which $\Delta = \Theta(\log n)$ and all blocks are good w.h.p., we recursively bound the number of bad blocks in each tessellation, concluding that at most a constant fraction of all blocks in the first tessellation are bad. Moreover, blocks that are far from each other by at least some constant distance (in space-time), satisfy the property of being good independently of one another. We can then use this result to show that any possible information path contains sufficiently many good blocks, which guarantees that information reaches from one end of the n-path to the other. We note that various aspects of our proof are simpler that in the original proof of Kesten and Sidoravicius, mainly because our process stores the information at vertices, resulting in information paths that are easier to analyse.

Road-map. In Sect. 2 we give an overview of the related work. In Sect. 3 we prove Theorem 2, and in Sect. 4 we prove the upper bound of Theorem 5. Due to space limitations, the proofs of the remaining results are only available in the full version of the paper [18].

2 Related Work

Independent parallel random walks have been studied since the late 70s [1], mainly as a way to speed-up cover and hitting times and related graph problems [2,4,7,8,12,13]. Similarly, randomized rumor spreading, where information exchange occurs between adjacent vertices (e.g., via push, pull, or push-pull), has been studied for the past 35 years [9,14,21], with the more recent results studying the spreading time in social networks [11], and bounds with graph expansion [5].

A closely related diffusion process to ours is the one where information is not stored on vertices, but is transmitted directly between agents when they meet, and initially a single agent is informed. Naturally, the spreading time in this setting is the time until all *agents* are informed. Several works have studied this process [8,10,16,26,28]. Dimitriou et al. [10], observed that on any graph the expected spreading time is $O(t^* \log m)$, where m is the number of agents (placed at arbitrary vertices, initially), and t^* is the maximum expected meeting time of two walks; this bound is tight for some graphs. Better bounds were also provided for the complete graph and expanders. Cooper et al. [8] showed (among other results) that the expected spreading time on a random d-regular graph converges to $\frac{2n \ln m}{m} \cdot \frac{d-1}{d-2}$, for most starting positions of the m agents. Pettarin et al. [28,29] considered the k-dimensional grid, $G_{k,n}$, for $k \in \{1,2\}$, and showed that the spreading time is $\tilde{\Theta}(n/\sqrt{m})$,[3] w.h.p., for m agents starting from stationarity. Lam et al. [26] studied the same problem for $k \geq 3$ dimensions, and showed a phase transition depending on m: for large m the spreading time is $\tilde{\Theta}(n^{1+1/k}/\sqrt{m})$, while for small m it is $\tilde{\Theta}(n/m)$. Giakkoupis et al. [16,17] considered the process on d-regular graphs, with $m = \Theta(n)$ agents starting from stationarity, and showed that, on any d-regular graph with $d = \Omega(\log n)$, the spreading time is asymptotically at least as large as for VISIT-EXCHANGE, and in some cases strictly larger.

Kesten and Sidoravicius [23,24] studied a continuous-time variant of the above process on the infinite grid, where the initial number of agents on each vertex is a poisson random variable with constant mean, and the information starts from the origin. They proved a theorem for the shape formed by the contour of the informed agents in the limiting case. In their analysis it is implicit that if the grid is finite, the spreading time is linear in the diameter (see also [19]). Our proof of Theorem 6 uses techniques from their analysis. A very similar process is the frog model, where only informed agents move, while uninformed ones stay at their initial position, until they are hit by an informed agent. At that point they get informed, and start their own walk. This process has been studied on infinite grids [3,30] and trees [20].

3 Upper Bound for Regular Graphs

In this section we prove Theorem 2.

3.1 Preliminaries

Let $G = (V, E)$ be any graph (not necessarily a regular one), and let A be the set of agents in VISIT-EXCHANGE, where $|A| = \alpha \cdot n$ for a constant $\alpha > 0$. The agents in A start their walks from the stationary distribution π. For a vertex u, let $N_u(t)$ be the number of agents

[3] The tilde asymptotic notation hides polylogarithmic factors.

that are at vertex u at round t. For an integer $r > 0$ and round t, let

$$\hat{N}_u(t,r) = \mathbb{E}\left[N_u(t+r) \mid N_v(t),\ \text{for all } v \in V\right] = \sum_{v\in V} p^r_{v,u} \cdot N_v(t),$$

where $p^r_{v,u}$ is the probability that a random walk starting from v is at u after exactly r rounds.

▶ **Lemma 7.** *For any vertex u, round t, and integer r,*

$$\mathbb{P}\left[\hat{N}_u(t,r) \le |A| \cdot \pi(u)/2\right] \le \exp\left(-\frac{|A| \cdot \pi(u)}{8 \cdot p^{2r}_{u,u}}\right).$$

Proof. Let $X^t_{v,g}$ be an indicator random variable, which is 1 when agent g is at vertex v at round t. Then, $N_v(t) = \sum_{g\in A} X^t_{v,g}$, which implies

$$\hat{N}_u(t,r) = \sum_{v\in V} p^r_{v,u} \sum_{g\in A} X^t_{v,g} = \sum_{g\in A}\sum_{v\in V} p^r_{v,u} \cdot X^t_{v,g} = \sum_{g\in A} Y_g,$$

where Y_g is the internal sum above for agent g. The random variables Y_g, $g \in A$, are independent, since the agents perform independent random walks. We compute the expectation and the second moment of Y_g to argue about the concentration of $\hat{N}_u(t,r)$.

$$\mathbb{E}\left[\hat{N}_u(t,r)\right] = \mathbb{E}\left[N_u(t+r)\right] = |A| \cdot \pi(u),$$

as the agents are initially distributed according to the stationary distribution π.

$$\begin{aligned}
\mathbb{E}\left[Y_g^2\right] &= \mathbb{E}\left[\sum_{v_1,v_2\in V} p^r_{v_1,u} p^r_{v_2,u} \cdot X^t_{v_1,g} \cdot X^t_{v_2,g}\right] \\
&= \sum_{v\in V} \left(p^r_{v,u}\right)^2 \cdot \mathbb{E}\left[X^t_{v,g}\right], \quad \text{as } g \text{ cannot be in two vertices simultaneously,} \\
&= \sum_{v\in V} p^r_{v,u} \cdot \left(p^r_{v,u} \cdot \pi(v)\right), \quad \text{since } g \text{ is placed according to } \pi, \\
&= \sum_{v\in V} p^r_{u,v} \cdot \left(\pi(u) \cdot p^r_{u,v}\right), \quad \text{by reversibility,} \\
&= \pi(u) \cdot \sum_{v\in V} p^r_{u,v} \cdot p^r_{v,u} \\
&= \pi(u) \cdot p^{2r}_{u,u}.
\end{aligned}$$

We apply [6, Theorem 3.7], setting $\lambda = \mathbb{E}\left[\hat{N}_u(t,r)\right]/2$ and $M = 0$, to obtain

$$\begin{aligned}
\mathbb{P}\left[\hat{N}_u(t,r) \le |A| \cdot \pi(u)/2\right] &\le \exp\left(-\frac{\lambda^2}{2 \cdot \sum_{g\in A} \mathbb{E}\left[Y_g^2\right]}\right) \\
&\le \exp\left(-\frac{(|A| \cdot \pi(u))^2}{8 \cdot \sum_{g\in A} \pi(u) \cdot p^{2r}_{u,u}}\right) = \exp\left(-\frac{|A| \cdot \pi(u)}{8 \cdot p^{2r}_{u,u}}\right). \quad ◀
\end{aligned}$$

We will also need the following result, whose proof is in the full version of the paper [18].

▶ **Lemma 8.** *Let $X(t)$ be a simple random walk that starts at vertex u of a connected graph $G = (V,E)$. If $\deg(u)$ is the degree of u, and $d_{\min}$ is the smallest degree of G, then for any even $t \ge 0$, $\mathbb{P}\left[X(t) = u\right] \le \frac{\deg(u)}{|E|} + \frac{20\cdot\deg(u)}{d_{\min}\cdot\sqrt{t+1}}$.*

3.2 Analysis

Suppose that $G = (V, E)$ is a d-regular graph with $d = O(\log n)$, thus $\pi(u) = 1/n$ for any $u \in V$. For a constant $\rho > 0$ define $r = r(\rho)$ as the smallest even integer such that

$$r \geq \max\{\rho \cdot \log^2 n,\ 256d \cdot \log n/\alpha\} = \Theta(\log^2 n). \tag{1}$$

We modify the VISIT-EXCHANGE process to create a new process called TWEAKED$_r$, as follows: At the end of each round $t \geq 0$, we add a minimal set of agents to the process to make sure that $\hat{N}_u(t, r) \geq |A| \cdot \pi(u)/2 = \alpha/2$, for every vertex u. Next we prove that, in the first polynomially many rounds TWEAKED$_r$ and VISIT-EXCHANGE are equivalent, w.h.p. Therefore, the results that we prove for TWEAKED$_r$, also hold for VISIT-EXCHANGE, w.h.p. This technique allows us to avoid dealing with dependencies of the random walks, which would arise if we directly analyzed VISIT-EXCHANGE conditioned on $\hat{N}_u(t, r) \geq \alpha/2$ for all u and t. (Similar tweaked processes are used in the proofs of Theorems 2 to 5 to circumvent some dependencies.)

▶ **Lemma 9.** *For any constant $c > 0$, there is a constant ρ such that VISIT-EXCHANGE and TWEAKED$_r$ are identical for the first T' rounds of their execution with probability at least $1 - T' \cdot n^{-(c+2)}$.*

Proof. By Lemma 8, $p^{2r}_{u,u} \leq \frac{2}{n} + \frac{20}{\sqrt{2r+1}} \leq \frac{20}{\sqrt{r}}$, since $r = O(\log^2 n)$. For $t < T'$, we substitute the above inequality into Lemma 7, and use the fact that $|A| \cdot \pi(u) = \alpha$, to get that

$$\mathbb{P}\left[\hat{N}_u(t, r) \leq \alpha/2\right] \leq \exp\left(-\frac{\alpha}{8 \cdot p^{2r}_{u,u}}\right) \leq \exp\left(-\frac{\alpha}{160} \cdot \sqrt{r}\right) \leq n^{-(c+3)},$$

for a sufficiently large constant ρ. By applying a union bound over all vertices u and rounds $t < T'$, we complete the proof. ◀

Consider two vertices u and v with distance $O(r/\max\{d, \log^2 \log n\})$, and assume u is informed at round t_0. The next key lemma provides a lower bound for the probability that v becomes informed $O(r)$ rounds after t_0. The lemma holds for any execution prefix of TWEAKED$_r$ up to round t_0, which means we can apply it repeatedly to prove Theorem 2. Let $\mathcal{K}_t$ be the σ-field that determines the execution of TWEAKED$_r$ until round t.

▶ **Lemma 10.** *Let $h = \max\{d, \log\log n\}$, and $k_{\max}(\gamma) = \frac{\gamma \cdot r}{\max\{d, (\log\log n)^2\}}$. There are constants $\gamma, \beta > 0$, such that for any round t_0 and any two vertices u, v with $\mathrm{dist}(u, v) \leq k_{max}(\gamma)$, given $\mathcal{K}_{t_0}$ and that u is informed at round t_0, vertex v is informed at round $t_0 + 2r$ with probability at least $1 - e^{-\beta \cdot h}$.*

Proof. Case $d = \omega(\log\log n)$. To simplify presentation, we assume $t_0 = 0$ and omit the conditional $\mathcal{K}_{t_0}$ throughout the proof. Fix the constant γ such that $k_{\max}(\gamma) \leq \frac{\alpha r}{256d}$. Consider two vertices u, v such that a shortest path between them is $u = u_0, \ldots, u_k = v$, where $k = \mathrm{dist}(u, v) \leq k_{\max}(\gamma)$. For a round $t \geq r$ and $i \in \{0, \ldots, k-1\}$, let $Z_{i,t}$ be the number of agents in the neighbourhood $\Gamma(u_i)$ of vertex u_i at round t. Then, by definition of TWEAKED$_r$,

$$\mathbb{E}\left[Z_{i,t}\right] = \sum_{w \in \Gamma(u_i)} \mathbb{E}\left[N_{u_i}(t)\right] = \sum_{w \in \Gamma(u_i)} \mathbb{E}\left[\hat{N}_{u_i}(t - r, r)\right] \geq \alpha \cdot d/2.$$

Since the agents make independent random walks, by a Chernoff bound we get that

$$\mathbb{P}\left[Z_{i,t} \geq \alpha \cdot d/4\right] \geq 1 - e^{-\alpha \cdot d/16}.$$

If $\mathcal{E}$ is the event that $Z_{i,t} \geq \alpha \cdot d/4$ for all $i \in \{0, \dots k-1\}$ and $t \in \{r, \dots 2r\}$ simultaneously, then, by a union bound,

$$\mathbb{P}[\mathcal{E}] \geq 1 - k \cdot r \cdot e^{-\alpha \cdot d/16} \geq 1 - e^{-\beta d}/2,$$

for a small enough constant β, because $kr = O(\text{poly}(\log n))$ and $d = \omega(\log \log n)$.

We modify TWEAKED$_r$ as follows: If $\mathcal{E}$ does not hold, then we add a minimum number of agents to the process so that $\mathcal{E}$ holds. We call the new process R-TWEAKED$_r$, and observe that TWEAKED$_r$ and R-TWEAKED$_r$ are identical with probability at least $1 - e^{-\beta d}/2$.

We divide the rounds $r, \dots, 2r-1$ of R-TWEAKED$_r$ into $r/2$ phases of 2 rounds each. For each $0 \leq i < r/2$, let $\mathcal{K}'_i$ be the σ-algebra which determines the execution prefix of R-TWEAKED$_r$ until round $r + 2i \leq 2r$. Let p_i be the largest integer, between 0 and k, such that vertex $w = u_{p_i}$ is informed at round $r + 2i$. If $p_i < k$, then each agent that is in the neighbourhood of w in round $r + 2i$, informs vertex u_{p_i+1} after two rounds, with probability $1/d^2$, by going through w. Define a Bernoulli random variable X_i, such that $X_i = 1$ if $p_i < k$ and u_{p_i+1} is informed in round $r + 2(i+1)$, i.e., the ith phase is successful. For technical convenience, we also define $X_i = 1$ if $p_i = k$, i.e., v is already informed in that phase. Then,

$$\mathbb{P}[X_i = 1 \mid \mathcal{K}'_i] \geq 1 - \left(1 - d^{-2}\right)^{\alpha \cdot d/4} \geq 1 - e^{-\alpha/(4d)} \geq \frac{\alpha}{8d}. \tag{2}$$

Define $Y = \sum_{i=0}^{r/2-1} Y_i$, where Y_i are independent Bernoulli random variables with success probability $\alpha/8d$. By our choice of γ and (1),

$$\mathbb{E}[Y] = \frac{\alpha r}{16d} \geq 8(k_{\max}(\gamma) + \log n) \geq 8(k + \log n),$$

and, by a Chernoff bound,

$$\mathbb{P}[Y \geq k] \geq \mathbb{P}[Y \geq \mathbb{E}[Y]/2] \geq 1 - e^{-\mathbb{E}[Y]/8} \geq 1 - 1/n \geq 1 - e^{-\beta d}/2,$$

since $d = O(\log n)$ and by choosing constant β smaller if necessary. On the other hand, for $X = \sum_{i=1}^{r/2-1} X_i$, (2) implies that X stochastically dominates Y, in particular,

$$\mathbb{P}[X \geq k] \geq \mathbb{P}[Y \geq k] \geq 1 - e^{-\beta d}/2.$$

Note, $X \geq k$ implies that v is informed in R-TWEAKED$_r$ at round $2r$. Since R-TWEAKED$_r$ and TWEAKED$_r$ are identical with probability $1 - e^{-\beta d}/2$, vertex v must be informed in TWEAKED$_r$ at round $2r$ with probability at least $1 - e^{-\beta d} = 1 - e^{-\beta h}$.

Case $d = O(\log \log n)$**.** As in the previous case, we assume $t_0 = 0$ and consider the spread of information along a shortest path from u to v, namely, $u = u_0, \dots, u_k = v$. Fix a round $t \geq r$ and some $i \in \{0, \dots, k-1\}$. Let $l = (\eta \log \log n)^2$ for some constant η that will be specified later. For an agent g define R_g as the number of times agent g visits u_i in rounds $t, \dots, t+l-1$. If $X_g(t')$ is the position of the agent g at round t', then $R_g = \sum_{t'=t}^{t+l-1} \mathbb{1}_{X_g(t')=u_i}$, so by Lemma 8,

$$\mathbb{E}[R_g \mid X_g > 0] = \sum_{t'=t}^{t+l-1} \mathbb{P}[X_g(t') = u_i \mid R_g > 0] \leq 1 + \sum_{t'=t}^{t+l-1} \left(\frac{1}{n} + \frac{20}{\sqrt{t - t' + 1}}\right) \leq 50 \cdot \sqrt{l}.$$

Let $Z_{i,t}$ be the number of unique agents that visit u_i in rounds $t, \dots, t+l-1$.

$$\mathbb{E}[Z_{i,t}] = \sum_{g \in A} \mathbb{P}[R_g > 0] = \sum_{g \in A} \frac{\mathbb{E}[R_g]}{\mathbb{E}[R_g \mid R_g > 0]}$$

$$\geq \frac{\sum_{g\in A}\mathbb{E}\left[R_g\right]}{50\cdot\sqrt{l}} = \frac{\sum_{t'=t}^{t+l-1}\mathbb{E}\left[N_{u_i}(t')\right]}{50\cdot\sqrt{l}} = \frac{\sum_{t'=t}^{t+l-1}\mathbb{E}\left[\hat{N}_{u_i}(t'-r,r)\right]}{50\cdot\sqrt{l}}$$
$$\geq \frac{l\cdot\alpha/2}{50\cdot\sqrt{l}} = \frac{\alpha\cdot\sqrt{l}}{100}.$$

Since the agents are performing independent random walks, then by a Chernoff bound,

$$\mathbb{P}\left[Z_{i,t}\geq\alpha\cdot\sqrt{l}/200\right]\geq 1-\exp\left(-\frac{\alpha\eta}{800}\cdot\log\log n\right)\geq 1-1/\log^5 n,$$

for a suitable choice of η. We now let $\mathcal{E}$ be the event $Z_{i,t}\geq\alpha\cdot\sqrt{l}/200$ for all $i\in\{0,\ldots,k-1\}$ and $t\in\{r,\ldots,2r\}$, simultaneously. As before, we create R-TWEAKED$_r$ by adding minimum number of agents to TWEAKED$_r$ to ensure that $\mathcal{E}$ holds. Since $rk=O(\log^4 n)$, by a union bound, there is a constant β such that $\mathbb{P}\left[\mathcal{E}\right]\geq 1-e^{-\beta h}/2$.

The rest of the proof follows the same line of logic as in the case of $d=\omega(\log\log n)$. The only difference is that instead of phases of 2 rounds, we consider phases of l rounds. $\mathcal{E}$ implies that after each phase R-TWEAKED$_r$ informs the next vertex on the path with a constant probability since $\sqrt{l}=\Omega(d)$. Therefore, as long as $k\leq\gamma\cdot r/l$ for a sufficiently small γ, vertex v becomes informed at round $2r$ of R-TWEAKED$_r$ w.h.p., which completes the proof. ◀

Proof of Theorem 2. First, we consider the TWEAKED$_r$ process for a constant ρ chosen by Lemma 9 such that TWEAKED$_r$ is identical to VISIT-EXCHANGE in the first n^2 rounds of its execution, with probability at least $1-n^{-2}$. Consider a shortest path $s=u_0,\ldots,u_m=u$ from source vertex s to vertex u. Let $k=k_{max}(\gamma)$ be the upper bound on the distance from Lemma 10, and as before $h=\max\{d,\log\log n\}$. We divide the execution of TWEAKED$_r$ into phases of $2r$ rounds each. If vertex u_i is informed at the end of a phase, then by Lemma 10, the vertex $u_{\min\{m,i+k\}}$ will be informed in the next phase of $2r$ rounds with probability at least $1-e^{-\beta h}$, independently from the past.

For some constant $\eta\in(0,1)$, let $l=\lceil m/k+\log n/h\rceil/(1-\eta)$. For $i\in\{1,\ldots,l\}$, let X_i be a Bernoulli random variable that is 0 if in the ith phase of TWEAKED$_r$ either k new vertices along the specified path become informed, or vertex u becomes informed, i.e., the phase is successful. For $X=\sum_{i=1}^{l}X_i$, if $X<l-\lceil m/k\rceil$ then vertex u is informed at the end of the lth phase, because at least $\lceil m/k\rceil$ phases were successful. By a stochastic dominance argument as in Lemma 10 we upper bound $\mathbb{P}\left[X<l-\lceil m/k\rceil\right]$.

Let $\{Y_i\}_{1\leq i\leq l}$ be a collection of independent Bernoulli random variables $\mathbb{P}\left[Y_i=1\right]=e^{-\beta h}$. By Lemma 10, $\mathbb{P}\left[X_i=1\mid X_1,\ldots,X_{i-1}\right]\leq\mathbb{P}\left[Y_i=1\right]$, and therefore, for $Y=\sum_{i=1}^{l}Y_i$,

$$\begin{aligned}\mathbb{P}\left[X>l-\lceil m/k\rceil\right]&\leq\mathbb{P}\left[Y>l-\lceil m/k\rceil\right]\leq\mathbb{P}\left[Y\geq l-\lceil m/k+\log n/h\rceil\right]\\&=\mathbb{P}\left[Y\geq\eta\cdot l\right]=\mathbb{P}\left[Y\geq\eta\cdot e^{\beta h}\cdot\mathbb{E}\left[Y\right]\right]\\&\leq\left(\eta\cdot e^{\beta h-1}\right)^{-\eta\cdot l}\leq n^{-3},\end{aligned}$$

by a Chernoff bound and by taking a value of η that is sufficiently close to 1. Thus, after $l\cdot 2r$ rounds of TWEAKED$_r$ vertex u is informed with probability $1-n^{-3}$. By a union bound over all vertices, and the fact that TWEAKED$_r$ and VISIT-EXCHANGE are identical in the first n^2 rounds we get that $T\leq l\cdot 2r$ w.h.p. Since $k=O(r/\max\{d,(\log\log n)^2\})$, and $m\leq\operatorname{diam}(G)$, and $h=\max\{d,\log\log n\}$, we finally get that, w.h.p.,

$$T=O\left(\max\{d,(\log\log n)^2\}\cdot\operatorname{diam}(G)+\frac{\log^3 n}{h}\right)=\tilde{O}\left(d\cdot\operatorname{diam}(G)+\frac{\log^3 n}{d}\right).$$ ◀

4 Upper Bound for Trees

In this section we prove the upper bound part of Theorem 5. Recall, $R_{b,h}$ is a rooted b-ary tree, where each vertex at distance less than h from the root has b children, and all leaves are at distance h from the root; thus h is the *height* of the tree. The total number of vertices is $n = (b^{h+1} - 1)/(b - 1)$. The set of children of vertex u is denoted C_u. The set of descendants of u is denoted D_u; precisely, D_u contains the vertices in the subtree rooted u, including u itself. The height of that subtree is denoted h_u. We define the set $B_{u,l} = \{v \in D_u \mid h_v = h_u - l\}$, which contains all descendants of v at distance l from u. Finally, $Z_u(t)$ denotes the set of agents at vertex u at round t, and $Z_S(t) = \bigcup_{u\in S} Z(t)$ is the set of agents in the set $S \subseteq V$ at that round.

4.1 The Lucky-Gambler Process

We define an auxiliary process, called LUCKY-GAMBLER, which will be used in the analysis. The process has three parameters: two integers $m, k > 0$, and a probability $p < 1/2$. Consider a path graph P_m of length m, with vertices 0 up to m. For every integer $s \geq 0$, at round s exactly k *gamblers* appear on vertex 1 and make a biased random walk: for $0 < i < m$, the probability of moving from vertex i to $(i + 1)$ and $(i - 1)$ is $p_{i,i+1} = p$ and $p_{i,i-1} = 1 - p = q$, respectively. When the gambler reaches vertex 0 or m, it stops, i.e., $p_{0,0} = p_{m,m} = 1$ (states $0, m$ are absorbing). We will write LUCKY-GAMBLER(m, p, k) to explicitly state the parameters of the process.

For a vertex v of $R_{b,h}$, where $h_v \geq m$, we are going to couple the movement of the agents in part of the subtree of v, with the gamblers in LUCKY-GAMBLER. Using the coupling and the next lemmas, we argue that v receives agents at a constant rate. By carefully selecting the agents that are coupled, we can claim that agents arrive at constant rate to every vertex v on a given path to the root, independently for each vertex.

▶ **Lemma 11.** *If $p = 1/(b+1)$ and $k \geq \epsilon \cdot b^{m-1}$, for some constant $\epsilon > 0$, then there is a constant $\beta < 1$ such that for any round $t \geq 4m$ and positive integer Δ the probability that no gambler reaches vertex m during any round in $\gamma_0 = \{t, \dots, t + \Delta - 1\}$ is at most $(1 - \beta)^{\Delta}$.*

▶ **Lemma 12.** *If $p = 1/(b+1)$ and $k \geq \kappa \cdot b^{m-1}$, for some integer κ, then there is a constant γ, such that for any integer $\tau \geq 8m$, at least $\gamma\kappa\tau$ gamblers reach vertex m in the first τ rounds, with probability at least $1 - e^{-\gamma\kappa\tau/4}$.*

We will use the next two results for a single gambler g making a biased random walk on P_m starting at round 0. Let $X_g(t)$ be the position of gambler g at round t and let $\tau_g(i) = \min\{t \mid X_g(t) = i\}$ be the hitting time of vertex i of g. We denote the event that $\tau_g(m) < \tau_g(0)$ as $\mathcal{L}_g$, and we will say that g is *lucky* if it occurs.

▶ **Lemma 13** ([15, Chapter 14]). *If $p \neq q$, then for $0 < i < m$, $\mathbb{P}[\mathcal{L}_g \mid X_g(0) = i] = \frac{(q/p)^i - 1}{(q/p)^m - 1}$.*

▶ **Lemma 14.** *If $p < q$, then for $0 < i < m$, $\mathbb{E}[\tau_g(m) \mid \mathcal{L}_g, X_g(0) = i] \leq \frac{m-i}{q-p}$.*

Below we prove Lemma 11; the proofs of Lemmas 12 and 14 can be found in the full version of the paper [18].

Proof of Lemma 11. For $s \geq 0$ and $1 \leq i \leq k$, let $g_{s,i}$ be the ith gambler that starts its walk at round s at vertex 1. Let $\tau_{s,i} = \tau_{g_{s,i}}$ be defined as for the single gambler g above. Clearly, $\tau_{s,i}(j) - s$ and $\tau_g(j)$ are identically distributed, if $X_g(0) = 1$. We also extend the definition of γ_0, letting $\gamma_s = \{t - s, \dots, t + \Delta - s - 1\}$.

We would like to study the number of lucky gamblers that reach m at rounds in γ_0. Consider first a "toy" example, which assumes that for each s, exactly one gambler is lucky among the k gamblers that start their walk at round s. Suppose that g'_s is that lucky gambler. We study the expected number of these agents that reach m during the rounds in γ_0:

$$\mathbb{E}\left[\sum_{s\geq 0}\mathbf{1}_{\{\tau_{g'_s}(m)\in\gamma_0\}} \mid \mathcal{L}_{g'_s} \text{ for } s\geq 0\right] = \sum_{s=0}^{t+\Delta}\mathbb{P}\left[\tau_{g'_s}(m)\in\gamma_0 \mid \mathcal{L}_{g'_s}\right] = \sum_{s=0}^{t+\Delta}\mathbb{P}\left[\tau_g(m)\in\gamma_s \mid \mathcal{L}_g\right].$$

The setup in the "toy" example is unlikely to occur, however, we use it as a motivation to lower bound the last quantity, which will be used in the main part of the proof.

$$\begin{aligned}
\sum_{s=0}^{t+\Delta}\mathbb{P}\left[\tau_g(m)\in\gamma_s \mid \mathcal{L}_g\right] &= \sum_{l=0}^{\Delta-1}\sum_{\substack{0\leq s\leq t+\Delta\\ s\equiv l \pmod{\Delta}}}\mathbb{P}\left[\tau_g(m)\in\gamma_s \mid \mathcal{L}_g\right], \\
&\qquad \text{the inner sum is over every } \Delta\text{th summand,} \\
&\geq \sum_{l=0}^{\Delta-1}\mathbb{P}\left[\tau_g(m)< t \mid \mathcal{L}_g\right], \quad \text{by union of disjoint events,} \\
&= \Delta\cdot\mathbb{P}\left[\tau_g(m)< t \mid \mathcal{L}_g\right] \\
&\geq \Delta\cdot\left(1-\frac{\mathbb{E}\left[\tau_g(m) \mid \mathcal{L}_g\right]}{t}\right), \quad \text{by Markov's inequality,} \\
&\geq \Delta\cdot\left(1-\frac{m\cdot(b+1)}{t\cdot(b-1)}\right), \quad \text{by Lemma 14 as } q-p=\frac{b-1}{b+1}, \\
&\geq \Delta\cdot\left(1-\frac{b+1}{4(b-1)}\right), \quad \text{since } t\geq 4m, \\
&\geq \Delta/4.
\end{aligned}$$

We can now bound the probability that no agent visits vertex m between rounds t and $t+\Delta$:

$$\begin{aligned}
\mathbb{P}\left[\bigcap_{\substack{0\leq s\leq t+\Delta\\ 1\leq i\leq k}}\{\tau_{s,i}(m)\notin\gamma_0\}\right] &= \prod_{s=0}^{t+\Delta}\left(\mathbb{P}\left[\tau_{s,i}(m)\notin\gamma_0\right]\right)^k, \quad \text{by independence of the walks,} \\
&= \prod_{s=0}^{t+\Delta}\left(\mathbb{P}\left[\tau_g(m)\notin\gamma_s\right]\right)^k \\
&= \prod_{s=0}^{t+\Delta}\left(1-\mathbb{P}\left[\mathcal{L}_g\right]\cdot\mathbb{P}\left[\tau_g(m)\in\gamma_s \mid \mathcal{L}_g\right]\right)^k \\
&= \prod_{s=0}^{t+\Delta}\left(1-\frac{b-1}{b^m-1}\cdot\mathbb{P}\left[\tau_g(m)\in\gamma_s \mid \mathcal{L}_g\right]\right)^k, \text{ by Lemma 13,} \\
&\leq \prod_{s=0}^{t+\Delta}\exp\left(-\frac{k\cdot(b-1)}{b^m-1}\cdot\mathbb{P}\left[\tau_g(m)\in\gamma_s \mid \mathcal{L}_g\right]\right) \\
&\leq \exp\left(-\epsilon\cdot\frac{b^{m-1}(b-1)}{b^m-1}\cdot\sum_{s=0}^{t+\Delta}\mathbb{P}\left[\tau_g(m)\in\gamma_s \mid \mathcal{L}_g\right]\right) \\
&\leq \exp\left(-\frac{\epsilon\Delta}{8}\right), \quad \text{by the analysis of the toy example.} \qquad \blacktriangleleft
\end{aligned}$$

4.2 Analysis

We define another auxiliary process, called TWEAKED, which is a slight modification of the original VISIT-EXCHANGE process. Let m be the *smallest* integer such that $b^m \geq \mu \cdot \ln n$ for a constant μ to be defined later, and let $k = \lceil \alpha \cdot b^m/8 \rceil$. Consider a vertex u of the tree, such that $h_u \geq m$, and recall that $B_{u,m}$ is the set of descendants of u at distance m. Let v be one of the children of u and define $Z'_{u,v}(t)$ be the set of agents that are in $B_{u,m-1}$ at round t and were in $B_{u,m} \setminus B_{v,m-1}$ the round before, i.e., $Z'_{u,v}(t) = Z_{B_{u,m-1}}(t) \cap Z_{B_{u,m} \setminus B_{v,m-1}}(t-1)$. For a round $t \geq 0$ let $q_{u,v}(t)$ be the smallest non-negative integer for which

$$|Z'_{u,v}(t)| + q_{u,v}(t) \geq \left\lceil \frac{\alpha}{8} \cdot |B_{u,m}| \right\rceil = \left\lceil \frac{\alpha}{8} \cdot b^m \right\rceil = k.$$

To construct TWEAKED we add exactly $q_{u,v}(t)$ agents in $B_{u,m-1}$ at round t (it is not important to which vertices in $B_{u,m-1}$ these agents are added).

To motivate the construction of TWEAKED, consider a vertex u and its child v, such that $m \leq h_u < h$. In round t of TWEAKED, there are at least k agents at vertices in $B_{u,m} \setminus B_{v,m-1}$ (of height $h_u - m$) that move closer to u in the next round. This allows us to couple these agents to that of gamblers in a LUCKY-GAMBLER$(m+1, 1/(b+1), k)$ process, and use our results from Sect. 4.1 to show that agents arrive at the parent of u at a constant rate. A key insight is that by not considering agents that are in descendants of v, the same argument can be made for vertex v, independently of u, if $h_v \geq m$ too. By repeating this argument, we show that in $O(\log n)$ rounds all vertices of height at least m are informed once one such vertex is informed. TWEAKED and LUCKY-GAMBLER are also used to analyse the spread of the message in the vertices of height at most m.

Using a Chernoff bound we can show that TWEAKED and VISIT-EXCHANGE are equivalent in the first polynomially many rounds, w.h.p.

▶ **Lemma 15.** *The probability that no agent is added in the* TWEAKED *process in the first r rounds is at least $1 - r \cdot n^{-\frac{\alpha \cdot \mu}{32}+1}$.*

We will use the same notation for TWEAKED and VISIT-EXCHANGE processes.

▶ **Lemma 16.** *Let u be any vertex of the tree $R_{b,h}$ such that $h_u \geq m$. For any constant $c > 0$, if u is informed, then after $O(\log n)$ rounds of* TWEAKED *the root ρ of $R_{b,h}$ gets informed, with probability at least $1 - n^{-c}$.*

Proof. Consider the path $u = u_1, \ldots, u_l = \rho$ from u to the root of the tree. Due to the symmetry of the tree, we can assume that the path is the "leftmost" path of the tree, i.e., for any $i \geq 1$, u_{i-1} is the leftmost child of u_i (for consistency, we let u_0 be the leftmost child of u_1). Roughly speaking, we show that for any i, the number of rounds between two consecutive visits to u_i (by a certain subset of agent) follows a geometric distribution, independently of the other $u_{i'}$. To that end, we couple the movement of agents of TWEAKED to $l-1$ independent instances of process LUCKY-GAMBLER$(m+1, 1/(b+1), k)$, one corresponding to each of the vertices u_i for $1 \leq i < l$.

Next we give some definitions and describe the coupling for a fixed i. For simplicity, define $B_i = B_{u_i,m}$ and $B'_i = B_i \setminus B_{u_{i-1},m-1} = \bigcup_{v \in C_{u_i} \setminus \{u_{i-1}\}} B_{v,m-1}$. I.e., B_i is the set of descendants of u_i at distance m from it, and to get B'_i we remove the descendants of u_{i-1} from B_i. Let $g_1, \ldots, g_{z_{i,t}}$ be the agents in TWEAKED that were at B'_i in round $t-1$ and moved closer to the root in the next round. By definition of TWEAKED, there are at least $k = \lceil \alpha \cdot b^m/8 \rceil$ such agents.

In the LUCKY-GAMBLER$(m+1, 1/(b+1), k)$ process that corresponds to vertex u_i, we start k gamblers in round t, denoted $g'_1, \ldots, g'_k$. For each $1 \leq j \leq k$, and for each round $t' \geq t$ until g_j reaches u_{i+1} or any vertex in B_i, the walks g_j and g'_j are coupled: if g_j moves closer to the root then g'_j moves to the right on the path, and if g_j moves away from the root, g'_j moves left. If g_j is at u_{i+1} or in $B_{u_i,m}$, then by the coupling, g'_j has finished its walk at one of the endpoints of the path. Before this happens we say that g_j is *i-coupled*.

Let $t_1 = 4 \cdot (m+1)$, and let t_{i+1} be the first round after t_i when u_{i+1} receives an i-coupled agent from u_i. Now, notice that by construction no agent can be i-coupled and i'-coupled at the same time for $i' \neq i$. It implies that the rounds when u_{i+1} receives i-coupled agents are independent from the walks of i'-coupled agents. On the other hand the walks of i-coupled agents are coupled with an independent LUCKY-GAMBLER process thus, Lemma 11 implies

$$\mathbb{P}\left[t_{i+1} - t_i \leq s \mid t_1, \ldots, t_i\right] = (1-\beta)^s = \mathbb{P}\left[F_i \geq s\right],$$

where $F_i \sim Geom(\beta)$, $1 \leq i < l$, are a collection of independent geometric random variables with success probability β. If τ_ρ is the round when the root is informed then $\tau_\rho \leq t_l = t_1 + \sum_{i=1}^{l-1}(t_{i+1} - t_i)$. It follows that $(\tau_\rho - t_1)$ is stochastically dominated by $F = \sum_{i=1}^{l-1} F_i$, and from a Chernoff bound for the sum of independent geometric random variables,

$$\mathbb{P}\left[\tau_\rho \geq f + t_1\right] \leq \mathbb{P}\left[F \geq f\right] \leq e^{-f \cdot \beta/8},$$

for any $f \geq 2h/\beta$. Since $t_1 = O(h)$, we can take a large enough $f = O(\log n)$, completing the proof. ◀

Next we prove that if vertex u of height $h_u = m$ is informed, then after at most $O(m \ln n)$ rounds a given leaf v in u's subtree becomes informed, w.h.p. For that, we first show that there are at least $\Theta(m \ln n)$ visits to u in those rounds (possibly multiple times by the same agent). Using a lower bound on the probability that an agent that is at u visits v before returning to u, we can show that one of these agents will visit v in $O(m \ln n)$ rounds, w.h.p.

▶ **Lemma 17.** *Let u be such that $h_u = m$. For any constant $c > 0$, there is a round $\tau = O(m \ln n)$ such that in the first τ rounds of TWEAKED, u is visited at least $c \cdot mb \cdot \ln n$ times, with probability at least $1 - n^{-cmb}$.*

Proof. For a round t, let $g_1, \ldots, g_{z_{u,t}}$ be the agents that are in $B_{u,m-1}$ at round t, and have also been at the leaf vertices $B_{u,m}$ in the previous round. By the definition of TWEAKED, $z_{u,t} \geq k$, where $k = \lceil \alpha b^m/8 \rceil$. We construct an instance of LUCKY-GAMBLER$(m, 1/(b+1), k)$ as follows. If $g'_1, \ldots, g'_k$ are the gamblers that started their walk at round t, then for each $1 \leq j \leq k$, the walk of agent g_j is coupled with the walk of the gambler g'_k: If g_j moves closer to the root of the tree, then g'_j moves right on the path and left otherwise. The coupling ends when g'_j arrives at either vertex 0 or m of its path. That corresponds to g_j either visiting a leaf vertex in $B_{u,m}$ or visiting vertex u.

Consider the first τ rounds of TWEAKED. Since $k \geq \alpha b^m/8$, we can apply Lemma 12 with parameter $\kappa = \alpha b/8$ to the coupled LUCKY-GAMBLER process. Let γ be the constant guaranteed by the lemma and let $\tau = \frac{8c}{\alpha\gamma} \cdot m \ln n$. Lemma 12 implies that in the first τ rounds of LUCKY-GAMBLER there are at least $\gamma\kappa\tau = c \cdot mb \cdot \ln n$ lucky gamblers, with probability at least $1 - e^{-\gamma\kappa\tau/4} = 1 - e^{-cmb \ln n} = 1 - n^{-cmb}$. Since each lucky gambler corresponds to a single visit to u by some agent, we complete the proof. ◀

▶ **Lemma 18.** *Let u be such that $h_u = m$ and let v be a leaf in the subtree of u. For any constant $c_l > 0$, if vertex u is informed then after at most $O(m \ln n)$ rounds of TWEAKED, vertex v is informed with probability at least $1 - n^{-c_l}$.*

Proof. Let τ be the round guaranteed by Lemma 17 for a constant $c > 0$. If after the first τ rounds of TWEAKED, there have been fewer than $cmb \ln n$ visits to u, then we add a minimal number of agents to u at round τ to have at least $cmb \ln n$ agents there. We call the resulting process TWEAKED_u. By Lemma 17 and an application of union bound over the first $\log^2 n = \omega(m \ln n)$ rounds, TWEAKED_u and TWEAKED are identical in the first $\Theta(m \ln n)$ rounds of execution with probability at least $1 - n^{-cmb} \log^2 n$. We therefore analyse TWEAKED_u.

For a round $t \le \tau$, consider an agent g that visits u at round t. Let $\mathcal{D}_{g,t}$ be the event that g moves to one of u's children at round $t+1$. Let also $\mathcal{E}_{g,t}$ be the event that g visits v before returning to u, and before round $\tau' = \tau + 8mb^{m-1}$. Clearly, $\mathcal{E}_{g,t}$ implies $\mathcal{D}_{g,t}$, and $\mathbb{P}[\mathcal{D}_{g,t}] = \frac{b}{b+1}$. Also, we can show that $\mathbb{P}[\mathcal{E}_{g,t} \mid \mathcal{D}_{g,t}] \ge 1/(12mb)$, by analysing a single random walk in $R_{b,m}$ that starts in the root of the tree [18]. Therefore,

$$\mathbb{P}[\mathcal{E}_{g,t}] = \mathbb{P}[\mathcal{E}_{g,t} \cap \mathcal{D}_{g,t}] = \mathbb{P}[\mathcal{D}_{g,t}] \cdot \mathbb{P}[\mathcal{E}_{g,t} \mid \mathcal{D}_{g,t}] \ge \frac{b}{b+1} \cdot \frac{1}{12mb} \ge \frac{1}{18mb}.$$

The probability that v is not visited by any informed agent before round τ' is at most

$$\mathbb{P}\left[\bigcap_{t \le \tau,\, g \in Z_u(t)} \neg \mathcal{E}_{g,t}\right] \le \left(1 - \frac{1}{18mb}\right)^{cmb \ln n} \le e^{-c \ln n/18} \le n^{-c/18} \le n^{-c_l - 1},$$

for a large enough constant c. Notice that $\tau' = \tau + 8mb^{m-1} = O(m \ln n)$ by the definition of m. Since TWEAKED and TWEAKED_u are identical in the first $\log^2 n$ rounds with probability at least $1 - n^{-cmb} \log^2 n$, v will be informed in $O(m \ln n)$ rounds in TWEAKED, with probability at least $1 - n^{-c_l - 1} - n^{-cmb} \log^2 n \ge 1 - n^{-c_l}$. ◀

Proof of the Upper Bound of Theorem 5. We will use the following simple symmetry lemma, which holds for any graph: If $T_{u,v}$ is the number of rounds of VISIT-EXCHANGE until vertex v is informed when the information originates at u, then the random variables $T_{u,v}$ and $T_{v,u}$ have the same distribution [18].

Consider the TWEAKED process, and suppose that the source of the information is vertex u with $h_u = m$, for m as defined at the beginning of Sect. 4.2. By Lemma 16, for an arbitrary constant c, there is $T_1 = O(\log n)$ such that the root ρ is informed by time T_1, with probability at least $1 - n^{-c}$. Lemma 15 then implies that the same bound T_1 holds for the VISIT-EXCHANGE process, with probability $p \ge 1 - n^{-c} - n^{-\alpha\mu/32}$, for an arbitrary large μ. From the symmetry lemma above, it follows that if ρ is the initial source of the information instead, then u becomes informed within T_1 rounds of VISIT-EXCHANGE with the same probability $p \ge 1 - n^{-c} - n^{-\alpha\mu/32}$.

Suppose again that information originates at some u with $h_u = m$, and let v be any leaf that is a descendant of u. From Lemma 18 and Lemma 15, for an arbitrary constant c, there is some $T_2 = O(m \log n)$, such that v gets informed after at most T_2 rounds of VISIT-EXCHANGE, with probability at least $1 - n^{-c} - n^{-\alpha\mu/32}$.

Combining the above we obtain that if ρ is the source of the information, then any given leaf v is informed after at most $T_1 + T_2$ rounds of VISIT-EXCHANGE, with probability at least $1 - 2n^{-c} - 2n^{-\alpha\mu/32}$. And by a union bound, all leaves (and thus all vertices) are informed within $T_1 + T_2$ rounds with probability at least $1 - 2n^{-c+1} - 2n^{-\alpha\mu/32+1}$.

Finally, by employing the symmetry argument above again, we obtain that for any source vertex (not just ρ), all vertices are informed within $2(T_1 + T_2)$ rounds with probability at least $1 - 4n^{-c+1} - 4n^{-\alpha\mu/32+1}$. Since $T_1 + T_2 = O(\log n + m \log n) = O(\log n + \log_b \log n \cdot \log n) = O(\log n + h \log h)$, the theorem follows. ◀

References

1. Romas Aleliunas, Richard M. Karp, Richard J. Lipton, László Lovász, and Charles Rackoff. Random walks, universal traversal sequences, and the complexity of maze problems. In *Proc. 20th IEEE Symposium on Foundations of Computer Science, FOCS*, pages 218–223, 1979. doi:10.1109/SFCS.1979.34.
2. Noga Alon, Chen Avin, Michal Koucký, Gady Kozma, Zvi Lotker, and Mark R. Tuttle. Many random walks are faster than one. *Combinatorics, Probability & Computing*, 20(4):481–502, 2011. doi:10.1017/S0963548311000125.
3. O. S. M. Alves, F. P. Machado, and S. Yu. Popov. The shape theorem for the frog model. *The Annals of Applied Probability*, 12(2):533–546, May 2002. doi:10.1214/aoap/1026915614.
4. Andrei Z. Broder, Anna R. Karlin, Prabhakar Raghavan, and Eli Upfal. Trading space for time in undirected s-t connectivity. *SIAM J. Comput.*, 23(2):324–334, 1994. doi:10.1137/S0097539790190144.
5. Flavio Chierichetti, George Giakkoupis, Silvio Lattanzi, and Alessandro Panconesi. Rumor spreading and conductance. *J. ACM*, 65(4):17:1–17:21, 2018. doi:10.1145/3173043.
6. Fan R. K. Chung and Lincoln Lu. Concentration inequalities and martingale inequalities: A survey. *Internet Mathematics*, 3(1):79–127, 2006. doi:10.1080/15427951.2006.10129115.
7. Colin Cooper. Random walks, interacting particles, dynamic networks: Randomness can be helpful. In *Proc. 18th International Colloquium on Structural Information and Communication Complexity, SIROCCO*, pages 1–14, 2011. doi:10.1007/978-3-642-22212-2_1.
8. Colin Cooper, Alan M. Frieze, and Tomasz Radzik. Multiple random walks in random regular graphs. *SIAM J. Discrete Math.*, 23(4):1738–1761, 2009. doi:10.1137/080729542.
9. Alan J. Demers, Daniel H. Greene, Carl Hauser, Wes Irish, John Larson, Scott Shenker, Howard E. Sturgis, Daniel C. Swinehart, and Douglas B. Terry. Epidemic algorithms for replicated database maintenance. *Operating Systems Review*, 22(1):8–32, 1988. doi:10.1145/43921.43922.
10. Tassos Dimitriou, Sotiris E. Nikoletseas, and Paul G. Spirakis. The infection time of graphs. *Discrete Applied Mathematics*, 154(18):2577–2589, 2006. doi:10.1016/j.dam.2006.04.026.
11. Benjamin Doerr, Mahmoud Fouz, and Tobias Friedrich. Social networks spread rumors in sublogarithmic time. In *Proc. 43rd ACM Symposium on Theory of Computing, STOC*, pages 21–30, 2011. doi:10.1145/1993636.1993640.
12. Klim Efremenko and Omer Reingold. How well do random walks parallelize? In *Proc. Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM*, pages 476–489, 2009. doi:10.1007/978-3-642-03685-9_36.
13. Robert Elsässer and Thomas Sauerwald. Tight bounds for the cover time of multiple random walks. *Theor. Comput. Sci.*, 412(24):2623–2641, 2011. doi:10.1016/j.tcs.2010.08.010.
14. Uriel Feige, David Peleg, Prabhakar Raghavan, and Eli Upfal. Randomized broadcast in networks. *Random Struct. Algorithms*, 1(4):447–460, 1990. doi:10.1002/rsa.3240010406.
15. William Feller. *An Introduction to Probability Theory and its Applications. Vol. 1.* Wiley series in probability and mathematical statistics. Wiley, 1968.
16. George Giakkoupis, Frederik Mallmann-Trenn, and Hayk Saribekyan. How to spread a rumor: Call your neighbors or take a walk? In *Proc. 38th ACM Symposium on Principles of Distributed Computing, PODC*, pages 24–33, 2019. doi:10.1145/3293611.3331622.
17. George Giakkoupis, Frederik Mallmann-Trenn, and Hayk Saribekyan. How to spread a rumor: Call your neighbors or take a walk? *CoRR*, abs/2006.02368, 2020. URL: https://arxiv.org/abs/2006.02368, arXiv:2006.02368.
18. George Giakkoupis, Hayk Saribekyan, and Thomas Sauerwald. Spread of information and diseases via random walks in sparse graphs. Research report, Inria, August 2020. URL: https://hal.inria.fr/hal-02913942.
19. Peter Gracar and Alexandre Stauffer. Percolation of lipschitz surface and tight bounds on the spread of information among mobile agents. In *Proc. Approximation, Randomization, and*

Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM, volume 116 of *LIPIcs*, pages 39:1–39:17, 2018. doi:10.4230/LIPIcs.APPROX-RANDOM.2018.39.

20 Jonathan Hermon. Frogs on trees? *Electron. J. Probab.*, 23:40 pp., 2018. doi:10.1214/18-EJP144.

21 Richard M. Karp, Christian Schindelhauer, Scott Shenker, and Berthold Vöcking. Randomized rumor spreading. In *Proc. 41st IEEE Symposium on Foundations of Computer Science, FOCS*, pages 565–574, 2000. doi:10.1109/SFCS.2000.892324.

22 Harry Kesten and Vladas Sidoravicius. Branching random walk with catalysts. *Electronic Journal of Probability*, 8(0), 2003. URL: https://projecteuclid.org/euclid.ejp/1464037578, doi:10.1214/EJP.v8-127.

23 Harry Kesten and Vladas Sidoravicius. The spread of a rumor or infection in a moving population. *The Annals of Probability*, 33(6):2402–2462, Nov 2005. Zbl: 1111.60074. doi:10.1214/009117905000000413.

24 Harry Kesten and Vladas Sidoravicius. A shape theorem for the spread of an infection. *Annals of Mathematics*, 167(3):701–766, May 2008. doi:10.4007/annals.2008.167.701.

25 Axel Kramer, Ingeborg Schwebke, and Günter Kampf. How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. *BMC Infectious Diseases*, 6(1):130, Dec 2006. doi:10.1186/1471-2334-6-130.

26 Henry Lam, Zhenming Liu, Michael Mitzenmacher, Xiaorui Sun, and Yajun Wang. Information dissemination via random walks in d-dimensional space. In *Proc. 23rd ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 1612–1622, 2012. doi:10.1137/1.9781611973099.128.

27 Russell Lyons and Shayan Oveis Gharan. Sharp bounds on random walk eigenvalues via spectral embedding. *International Mathematics Research Notices*, 2018(24):7555–7605, 05 2017. doi:10.1093/imrn/rnx082.

28 Alberto Pettarin, Andrea Pietracaprina, Geppino Pucci, and Eli Upfal. Infectious random walks. *CoRR*, abs/1007.1604:21 pp., 2010. URL: http://arxiv.org/abs/1007.1604, arXiv:1007.1604.

29 Alberto Pettarin, Andrea Pietracaprina, Geppino Pucci, and Eli Upfal. Tight bounds on information dissemination in sparse mobile networks. In *Proc. 30th ACM Symposium on Principles of Distributed Computing, PODC*, pages 355–362, 2011. doi:10.1145/1993806.1993882.

30 Serguei Yu. Popov. Frogs and some other interacting random walks models. In *Proc. Discrete Random Walks, DRW*, pages 277–288, 2003. URL: http://dmtcs.episciences.org/3328.