Information Dissemination via Random Walks in d-Dimensional Space

Henry Lam * Zhenming Liu † Michael Mitzenmacher ‡ Xiaorui Sun § Yajun Wang ¶

Abstract

We study a natural information dissemination problem for multiple mobile agents in a bounded Euclidean space. Agents are placed uniformly at random in the d-dimensional space $\{-n, ..., n\}^d$ at time zero, and one of the agents holds a piece of information to be disseminated. All the agents then perform independent random walks over the space, and the information is transmitted from one agent to another if the two agents are sufficiently close. We wish to bound the total time before all agents receive the information (with high probability). Our work extends Pettarin et al's work [10], which solved the problem for $d \leq 2$. We present tight bounds up to polylogarithmic factors for the case $d = 3$. (While our results extend to higher dimensions, for space and readability considerations we provide only the case $d = 3$ here.) Our results show the behavior when $d \geq 3$ is qualitatively different from the case $d \leq 2$. In particular, as the ratio between the volume of the space and the number of agents varies, we show an interesting phase transition for three dimensions that does not occur in one or two dimensions.

1 Introduction

We study the following information diffusion problem: let $\mathrm{a}_1, \mathrm{a}_2, ..., \mathrm{a}_m$ be m agents initially starting at locations chosen uniformly at random in $\mathcal{V}^d = \{-n, -(n-1), \ldots, n\}^d$ and performing independent random walks over this space. One of the agents initially has a message, and the message is transmitted from one agent to another when they are sufficiently close. We are interested in the time needed to flood the message, that is, the time when all agents obtain the message. In other settings, this problem has been described as a virus diffusion problem, where the message is replaced by a virus that spreads according to proximity. We use *information diffusion* and *virus spreading* interchangeably, depending on which is more useful in context. This is a natural model that has been extensively studied. For example, Alves et al. and Kesten et al. coined the name "frog model" for this problem in the virus setting, and studied the shape formed by the infected contour in the limiting case [1, 2, 7]. In the flooding time setting, early works used a heuristic approximation based on simplifying assumptions to characterize the dynamics of the spread of the message [3, 8, 13]. More recent works provide fully rigorous treatments under this or similar random walk models [4, 5, 10, 12, 9].

The most relevant recent works are those of Pettarin et al. [10] and Peres et al. [9, 12]. The work of Pettarin et al. examines the same model as ours, but their analysis is only for one- and two-dimensional grids. The work of Sinclair and Stauffer [12] considers a similar model they call mobile geometric graphs, and their work extends to higher dimensions. However, their focus and model both have strong differences from ours. For example, they assume a Poisson point process of constant intensity, leading to a number of agents linear in the size of the space. In contrast, our results allow a sublinear number of agents, a scenario not directly relevant to their model. Also, they focus on structural aspects on the mobile graphs, such as percolation, while we are primarily interested in the diffusion time. There are additional smaller differences, but the main point is that for our problem we require and introduce new techniques and analysis.

Our paper presents matching lower bounds and upper bounds (up to polylogarithmic factors) for the flooding problems in d-dimensional space for an arbitrary constant d. For ease of exposition, in this paper we focus on the specific case where $d = 3$, which provides the main ideas. Two- and three-dimensional random walks have quite different behaviors – specifically, two-dimensional random walks are recurrent while three-dimensional random walks are transient – so it is not surprising that previous results for two dimensions fail to generalize immediately to three-dimensional space. Our technical contributions include new techniques and tools for tackling the flooding problem by building sharper approximations on the effect of agent interactions. The techniques developed in this paper

*Department of Mathematics and Statistics, Boston University. Email: khlam@math.bu.edu.

†Harvard School of Engineering and Applied Sciences. Supported by NSF grant CCF-0915922. Email: zliu@eecs.harvard.edu.

‡Harvard School of Engineering and Applied Sciences. Supported in part by NSF grants CCF-0915922 and IIS-0964473. Email: michaelm@eecs.harvard.edu.

§Columbia University. Email: xiaoruisun@cs.columbia.edu.

¶Microsoft Research Asia. Email: yajunw@microsoft.com.

are also robust enough that our results can be extended to variations of the model, such as allowing probabilistic infection rules, replacing discrete time random walks by continuous time Brownian motions, or allowing the agents to make jumps [5]. These extensions will be reported in future work.

Although the information diffusion problem in three or more dimensions appears less practically relevant than the two-dimensional case, we expect the model will still prove valuable. For instance, particles in a high dimensional space may provide a latent-space representation of the agents in a dynamic social network [6, 11], so understanding such information diffusion processes may be helpful for designing appropriate latent space models in the future. Also, the problem is mathematically interesting in its own right.

1.1 Our models and results We follow the model developed in [10]. Let $\mathcal{V}^d = \{-n, -(n-1), ..., 0, ..., (n-1), n\}^d$ be a d-dimensional grid. Let $A = \{\mathrm{a}_1, \mathrm{a}_2, ..., \mathrm{a}_m\}$ be a set of moving agents on $\mathcal{V}^d$. At $t = 0$, the agents spread over the space according to some distribution $\mathcal{D}$. Throughout this paper, we focus on the case where $\mathcal{D}$ is uniform. Agents move in discrete time steps. Every agent performs a symmetric random walk defined in the natural way. Specifically, at each time step an agent not at a boundary moves to one of its $2d$ neighbors, each with probability $1/(2d)$. If an agent is at a boundary, so there is no edge in one or more directions, we treat each missing edge as a self-loop. Let $\Xi_1(t), ..., \Xi_m(t) \in \{0, 1\}$ each be a random variable, where $\Xi_i(t)$ represents whether the agent a_i is infected at time step t. We assume $\Xi_1(0) = 1$ and $\Xi_i(0) = 0$ for all $i \neq 1$. The value $\Xi_i(t)$ will change from 0 to 1 if at time t it is within distance 1 to another infected agent a_j. (We use distance 1 instead of distance 0 to avoid parity issues.) Once a value $\Xi_j(t)$ becomes 1, it stays 1.

DEFINITION 1.1. (Information diffusion problem). *Let $A_1, A_2, \ldots, A_m \in \mathcal{V}^d$ be the initial positions of the agents $\mathrm{a}_1, \ldots, \mathrm{a}_m$ and let $S_t^1(A_1), S_t^2(A_2), \ldots, S_t^m(A_m)$ be m independent random walks starting at $A_1, \ldots, A_m$ respectively, so that $S_t^i(P)$ is the position of agent a_i at time t given that at $t = 0$ its position was $P \in \mathcal{V}^d$. The* infectious state *of each agent at time step t is a binary random variable $\Xi_i(t)$ such that*

- $\Xi_1(0) = 1$, $\Xi_i(0) = 0$ *for all other i, and*
- *for all $t > 0$, $\Xi_i(t) = 1$ if and only if*

$$(\Xi_i(t-1) = 1) \text{ or } \left(\exists j : \Xi_j(t-1) = 1 \wedge \left\|S_t^i(A_i) - S_t^j(A_j)\right\|_1 \leq 1\right).$$

The finishing time *of the diffusion process, or the* diffusion time, *is $T = \inf\{t \geq 0 : |\{\Xi_i(t) = 1\}| = m\}$.*

The following results for the diffusion time for one- and two-dimensional spaces are proved in [10].

THEOREM 1.1. *Consider the information diffusion problem for $d = 1, 2$ dimensions, and assume the agents are initially uniformly distributed over $\mathcal{V}^d$. Then, with high probability,*

$$T = \tilde{\Theta}(n^2 \cdot m^{-1/d}). \tag{1.1}$$

It is natural to ask whether Equation 1.1 also holds for $d \geq 3$. Our results show this is not the case.

THEOREM 1.2. (Diffusion time for $d \geq 3$) *Consider the information diffusion problem for $d \geq 3$ with initially uniformly distributed agents over $\mathcal{V}^d$. Then there exists a constant c such that*

$$\begin{aligned} &\textit{if} \quad cn^{d-2}\log^2 n < m < n^d : \\ &\qquad T = \tilde{\Theta}(n^{d/2+1} \cdot m^{-1/2}) \textit{ with high probability;} \\ &\textit{if} \quad m < cn^{d-2}\log^{-2} n : \\ &\qquad T \leq \tilde{\Theta}(n^d/m) \textit{ with high prob. and } T \geq \tilde{\Theta}(n^d/m) \textit{ a.s.} \end{aligned} \tag{1.2}$$

Notice that Theorems 1.2 and 1.1 yield the same result for $d = 2$, as well as when $d = 1$ and $m = \Theta(n)$. Here when we say with high probability, we mean the statement holds with probability $1 - n^{-\gamma}$ for any constant γ and suitably large n. When we say almost surely, we mean with probability $1 - o(1)$. When $m \geq n^d$, the result is implicit in [7] and the diffusion time in this case is $\tilde{\Theta}(n)$. Finally, there are some technical challenges regarding the case $cn^{d-2}\log^{-2} n \leq m \leq cn^{d-2}\log^2 n$ that we expect to address in future work.

An interesting point of our result is that when the number of agents m is greater than n^{d-2}, the finishing time is less than the mixing time of each individual random walk, and therefore the analysis requires techniques that do not directly utilize the mixing time. The rest of this extended abstract focuses on the lower and upper bounds for this interesting case, and, as previously mentioned, only for $d = 3$.

Theorem 1.2 can also be expressed in terms of the density of agents. Let $\lambda = m/n^d$ be the density. We can express the diffusion time as $T = \tilde{\Theta}(n/\sqrt{\lambda})$ w.h.p. for $cn^{-2}\log^2 n < \lambda < 1$, whereas for $\lambda < cn^{-2}\log^{-2} n$ we have $T \leq \tilde{\Theta}(1/\lambda)$ w.h.p. and $T \geq \tilde{\Theta}(1/\lambda)$ almost surely.

We remark that all theorems/propositions/lemmas in this paper are assumed to hold for sufficiently large n, but for conciseness we may not restate this condition in every instance.

2 Lower bound

Let us first state our lower bound result more precisely as follows.

THEOREM 2.1. *Let* $a_1, ..., a_m$ *be placed uniformly at random on* $\mathcal{V}^3$ *such that* $1600n\log^2 n \leq m \leq n^3$. *Let* $\ell_2 = \sqrt{n^3/m}$. *For sufficiently large* n, *the diffusion time* T *satisfies the following inequality*

$$\Pr[T \leq \frac{1}{81}\ell_2 n \log^{-29} n] \leq \exp(-\log n \log\log n).$$

We use a *local analysis* to prove our lower bound. The key idea is that under uniform distribution of agents, the extent any particular infected agent can spread the virus within a small time increment is confined to a small neighborhood with high probability. By gluing together these local estimates, we can approximate the total diffusion time. To explain our local analysis, assume we start with an arbitrary infected agent, say a_1. Let us also assume, for simplicity, that all the other uniformly distributed agents are uninfected. Consider the scenario within a small time increment, say Δt. During this time increment the agent a_1 infects whoever it meets in the small neighborhood that contains its extent of movement. The newly infected agents then continue to move and infect others. The size of the final region that contains all the infected agents at Δt then depends on the rate of transmission and the extent of movement of all of the infected agents. In particular, if Δt is small enough, the expected number of transmissions performed by a_1 is less than one; even if it infects another agent, the number of infections it causes within the *same* Δt is also less than one, and so on. The net effect is an eventual dying-out of this "branching process" (which we later model by what we call a diffusion tree), which localizes the positions of all infected agents at time Δt to a small neighborhood around the initial position of a_1.

Before laying out our analysis, let us briefly review the main methodologies in obtaining lower bound results in related work, and point out their relation to our analysis and difficulties in directly applying them to higher dimensions. Two potential existing methods arise in [2, 7] and [10]. The former analyzes the growth rate of the size of the *total* infected region; an upper bound on this growth rate translates to a lower bound for the diffusion time. The latter work, focusing on $d = 1, 2$, uses an "island diffusion rule", which essentially speeds up infection by allowing infections to occur immediately on connected components in an underlying graph where edges are based on the distance between agents. This approach avoids handling the issue of the meeting time of random walks when they are very close, a regime where standard asymptotic results may not apply, while still providing a way to bound the diffusion time by arguing about the low probability of interaction among different "islands".

The results in [2, 7] are not directly applicable in our setting because the growth rate they obtain is linear in time, as a result of their assumption of constant agent density in an infinite space, in contrast to our use of a size parameter n that scales with the agent density. It is fairly simple to see that blindly applying a linear growth rate to our setting of $o(1)$ density is too crude. On the other hand, analyzing how agent density affects the growth rate is a potentially feasible approach but certainly not straightforward.

Our approach more closely follows [10]. The main limitation of [10], when applied to higher dimensions, is how to control the interaction among islands. If islands interact too often, because they are too close together, the argument, which is based on a low probability of interaction, breaks down. However, if one parametrizes islands to prevent such interaction, then the bound that can be obtained is too weak. For $d > 2$ this constraint ultimately limits the analysis for the case of $o(1)$ density. We attempt to remedy the problem by using islands as an intermediate step to obtain local estimates of the influence of each initially infected agent over small periods of time. This analysis involves looking at a branching process representing the spread of the infection, significantly extending the approach of [10].

2.1 Local diffusion problem This subsection focuses on the local analysis as discussed above. In Section 2.2, we will proceed to discuss how to utilize this analysis to get the lower bound in Theorem 2.1. The two main difficulties in our analysis are: 1) probabilistic estimates for the meeting time/position of multiple random walks are typically only useful asymptotically; 2) walks near the boundary introduce further analytical complications. To begin with, the following definition serves to handle the second issue:

DEFINITION 2.1. (INTERIOR REGION) *The interior region* $\mathfrak{V}(r)$ *parameterized by* r *is the set of lattice points in* $\mathcal{V}^3$ *that have at least* L_∞*-distance* r *to the boundary.*

For any point $P \in \mathcal{V}^3$, define $\mathbb{B}(P, x) = \{Q \in \mathcal{V}^3 : \|Q - P\|_\infty \leq x\}$ as the x-ball of neighborhood of P under the L_∞-norm. We have the following result:

PROPOSITION 2.1. *Consider a diffusion following Definition 1.1. Let* S_0 *be the initial position of the only infected agent* a_1 *at time 0, and* $\mathcal{W}$ *be an arbitrary subset of lattice points in* $\mathfrak{V}(20\ell_2 \log n)$, *where* $\ell_2 = \sqrt{n^3/m}$. *Denote* $\Delta t = \ell_2^2 \log^{-28} n$. *Define the binary random variable* $b(\mathcal{W})$ *as follows:*

- If $S_0 \in \mathcal{W}$: $b(\mathcal{W})$ *is set as 1 if and only if all the*

infected agents at time Δt can be covered by the ball $\mathbb{B}(S_0, 9\ell_2 \log n)$.

- If $S_0 \notin \mathcal{W}$: $b(\mathcal{W}) = 1$.

We have $\Pr[b(\mathcal{W}) = 1] \geq 1 - \exp(-5 \log n \log\log n)$.

The proposition yields that with high probability, all the infected agents lie within a neighborhood of distance $\tilde{O}(\ell_2)$ at time $\tilde{O}(\ell_2^2)$. The variable ℓ_2 is chosen such that the expected number of infections spanned by an initially infected agent a_1 within $\tilde{O}(\ell_2^2)$ units of time and a neighborhood of $\tilde{O}(\ell_2)$ distance is $O(1)$. This can be seen by solving $m(\ell_2/n)^3 \times (1/\ell_2) = \tilde{O}(1)$, where $m(\ell_2/n)^3$ is the expected number of agents in a cube of size $\ell_2 \times \ell_2 \times \ell_2$, and $\tilde{O}(1/\ell_2)$ is the meeting probability within time $\tilde{O}(\ell_2^2)$ between any pair of random walks with initial distance ℓ_2. This choice of ℓ_2 appears to be the right threshold for our analysis. Indeed, a larger scale than ℓ_2 would induce a large number of infections made by a_1, and also subsequent infections made by newly infected agents, with an exploding affected region as an end result. On the other hand, a smaller scale than ℓ_2 would degrade our lower bound. This is because the diffusion time is approximately of order n/ℓ_2, the number of spatial steps to cover $\mathcal{V}^3$, times ℓ_2^2, the time taken for each step, equaling $n\ell_2$. Hence a decrease in ℓ_2 weakens the bound[1].

Secondly, we introduce $\mathcal{W}$ in Proposition 2.1 to avoid the case when S_0 is close to the boundary. As we have mentioned, such boundary conditions often complicate random walk analysis. Although the impact of the boundary's presence has been addressed in previous work (e.g., [5, 10]), existing results are not fully satisfactory. For example, when two simple random walks S^1 and S^2 start near the boundary, only a lower bound for the probability that two walks meet within a specific number of time steps is available ([10]); we do not know of an upper bound counterpart. We arrange our proof so that it is sufficient to analyze the diffusion pattern of a virus when it starts far from the boundary. Finally, we note that no effort has been made to optimize the exponent 28 in Δt's definition.

We briefly explain how our global lower bound can be readily obtained from Proposition 2.1, which is a strong characterization of the local growth rate of infection region size. Imagine the following evolution. Starting with a single infected agent, with high probability the infection spreads to a ball of radius at most $9\ell_2 \log n$ in Δt time units. At this time point, the newly infected agents *inside* the ball continue to spread the virus to neighborhoods of size at most $9\ell_2 \log n$, again with high probability. This gives an enlarged area of infection with radius at most $18\ell_2 \log n$. Continuing in this way, the lower bound in Theorem 1.2 is then the time for the infection to spread over $\mathcal{V}^3$. This observation will be made rigorous in the next subsection.

The rest of this subsection is devoted to the arguments of Proposition 2.1. It consists of two main steps. First, we need to estimate the expected number of infections done by a single initially infected agent within distance $9\ell_2 \log n$ and time increment Δt. Second, we iterate to consider each newly infected agent. The analysis requires the condition that the global configuration behaves "normally", a scenario that occurs with suitably high probability, as we show. We call this condition "good behavior", which is introduced through the several definitions below:

DEFINITION 2.2. (Island, [10]) *Let $A = \{\mathrm{a}_1, ..., \mathrm{a}_m\}$ be the set of agents in $\mathcal{V}^3$. For any positive integer $\gamma > 0$, let $G_t(\gamma)$ be the graph with vertex set A such that there is an edge between two vertices if and only if the corresponding agents are within distance γ (under L_1-norm) at time t. The island with parameter γ of an agent $\mathrm{a}_i \in A$ at time step t, denoted by $\mathrm{Isd}_t(\mathrm{a}_i, \gamma)$ is the connected component of $G_t(\gamma)$ containing a_i.*

DEFINITION 2.3. (GOOD BEHAVIOR) *Let $\ell_1 = nm^{-1/3}$. For $1 \leq i \leq (\ell_2/\ell_1)\log^{-3} n$, define $\mathcal{B}_i(P) = \mathbb{B}\left(P, i\ell_1 \log^{-1} n\right)$ and let $\partial\mathcal{B}_i(P) = \mathcal{B}_i(P) - \mathcal{B}_{i-1}(P)$. For any $P \in \mathcal{V}^3$, define $m_i(P) = \frac{(\log^5 n)|\partial\mathcal{B}_i(P)|m}{(2n+1)^3}$. Let us define the following binary random variables:*

- **Good density.** *Let $\{D_t : t \geq 0\}$ be a sequence of $0, 1$ random variables such that $D_t = 1$ if and only if for all $P \in \mathcal{V}^3$ and all $i \leq (\ell_2/\ell_1)\log^{-3} n$, the number of agents in $\partial\mathcal{B}_i(P)$ is at most $m_i(P)$, for all time steps up to t. We say the diffusion process has the* good density *property at time t if $D_t = 1$.*

- **Small islands.** *Let $\{E_t : t \geq 0\}$ be a sequence of $0, 1$ random variables such that $E_t = 1$ if and only if $|\mathrm{Isd}_s(\mathrm{a}_j, \ell_1 \log^{-1} n)| \leq 3 \log n$ for all $\mathrm{a}_j \in A$ and $0 \leq s \leq t$. We say that the diffusion process has the* small islands *property at time t if $E_t = 1$.*

- **Short travel distance.** *Let $\{L_t : t \geq 0\}$ be a sequence of $0, 1$ random variables such that $L_t = 1$ if and only if for all $i \in [m]$ and all $t_1 < t_2 \leq t$ with $t_2 - t_1 \leq \ell_2^2 \log^{-12} n$, we have $\|S^i_{t_1} - S^i_{t_2}\|_1 \leq 3\ell_2 \log^{-4} n$. We say the process has the* short travel distance *property at time t if $L_t = 1$.*

[1] In the case of general d-dimensional space, ℓ_2 is chosen such that $m(\ell_2/n)^d \times (1/\ell_2^{d-2}) = \tilde{O}(1)$, giving $\ell_2 = \sqrt{n^d/m}$. Throughout the paper such d-dimensional analog can be carried out in similar fashion, but for ease of exposition we shall not bring up these generalizations and will focus on the 3-dimensional case.

Finally, let $G_t = D_t \times E_t \times L_t$, *and say the diffusion process* behaves well *at time* t *if* $G_t = 1$. *We also focus on* $t \le n^{2.5}$ *and define the random variable* $G = G_{n^{2.5}}$.

The value $n^{2.5}$ in the definition is chosen such that it lies well beyond our lower bound for the case $m < n^3$, but is small enough for our forthcoming union bound. By using properties of random walks and techniques derived in [10], we have

LEMMA 2.1. *Let* $A = \{\mathrm{a}_1, ..., \mathrm{a}_m\}$ *be agents that are distributed uniformly in* $\mathcal{V}^3$ *at* $t = 0$. *For sufficiently large* n, *we have* $\Pr[G = 1] \ge 1 - \exp(-6 \log n \log\log n)$.

With this global "good behavior", we have the following estimate:

LEMMA 2.2. *Let* $A = \{\mathrm{a}_1, \ldots, \mathrm{a}_m\}$ *be agents that are distributed in* $\mathcal{V}^3$ *in such a way that* $D_0 = 1$. *Let* $S^1, S^2, \ldots, S^m$ *be their corresponding random walks. Consider an arbitrary agent* a_j *with* $S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n)$. *Let* $\{\mathrm{a}_{i_1}, \ldots, \mathrm{a}_{i_k}\}$ *be the set of agents outside* $\mathcal{B}_1(S_0^j)$ *at time* 0. *Define* $X_{j,\ell}$ *as the indicator random variable that represents whether the agents* a_j *and* a_{i_ℓ} *meet within time* $[0, \Delta t]$. *We have*

$$\mathrm{E}\left[\sum_{\ell \le k} X_{j,\ell} \middle| D_0 = 1, S_0^j \in \mathfrak{V}(2\ell_2 \log^{-4} n)\right] < \log^{-3} n.$$

Lemma 2.2 says that if the initial distribution of agents possesses good behavior, then the expected number of direct infections on far-away agents is small. For agents close to the initially infected agents, we instead utilize the concept of islands, and formally introduce a new diffusion process with a modified "island diffusion" rule. It is easy to see that the new diffusion process can be coupled with the original diffusion process (evolving with Definition 1.1) by using the same random walks in the same probability space.

DEFINITION 2.4. (Diffusion Process with Island Diffusion Rule) *Consider a diffusion process in which* m *agents are performing random walks on* $\mathcal{V}^3$. *An uninfected agent* a_j *becomes infected at time* t *if one of the following conditions holds:*

1. *it meets a previously infected agent at time* t. *For convenience, we say* a_j *is* directly *infected if it is infected in this way.*
2. *it is inside* $\mathrm{Isd}_t(\mathrm{a}_i, \ell_1 \log^{-1} n)$ *where* a_i *is directly infected at time* t.

We shall call the coupled process in Definition 2.4 the diffusion process *with island diffusion rule*. This process is different from the diffusion models introduced in [10, 12, 9]. In our formulation, an island is infected only if meeting occurs between one uninfected and one previously infected agent. In [10, 12, 9] (using our notations), an island is infected once it contains a previously infected agent. As a result, infections occur less frequently in our model than the models in [10, 12, 9]. This difference is the key to getting a tight lower bound for dimensions higher than 2. More precisely, our infection rule allows us to build a terminating branching process, or what we call a "diffusion tree" in the following definition, whose generations are defined via the infection paths from the source. The termination of this branching process constrains the region of infection to a small neighborhood around the source with a probability of larger order than obtained in [10]. This in turn leads to a tighter global lower bound.

DEFINITION 2.5. (DIFFUSION TREE) *Let* $\mathcal{W} \subseteq \mathfrak{V}(2\ell_2 \log n)$ *be a subset of lattice points. Consider a diffusion, following the island diffusion rule, that starts with an initially infected island* $\mathrm{Isd}_0(\mathrm{a}_1, \ell_1 \log^{-1} n)$. *Recall that* S_0^1 *denotes* a_1*'s position at* $t = 0$. *The diffusion tree* Tr *with respect to* $\mathcal{W}$ *has the following components:*

1. *If* $S_0^1 \notin \mathcal{W}$, $\mathrm{Tr} = \emptyset$.
2. *If* $S_0^1 \in \mathcal{W}$,
   - *The* root *of* Tr *is a dummy node* r.
   - *The children of* r *are all the agents in* $\mathrm{Isd}_0(\mathrm{a}_1, \ell_1 \log^{-1} n)$.
   - $\mathrm{a}_{\ell'}$ *is a* child *of* a_ℓ *(*$\mathrm{a}_{\ell'} \in \mathrm{child}(\mathrm{a}_\ell)$*) if* $\mathrm{a}_{\ell'}$ *is infected by* a_ℓ *before time* Δt.
   - $\mathrm{a}_{\ell'}$ *is a* direct child *of* a_ℓ *(*$\mathrm{a}_{\ell'} \in \mathrm{dchild}(\mathrm{a}_\ell)$*) if* $\mathrm{a}_{\ell'} \in \mathrm{child}(\mathrm{a}_\ell)$ *and it is* directly *infected by* a_ℓ.

For technical reasons, if $\mathrm{a}_{\ell'}$ *is not in* Tr, *we let* $\mathrm{child}(\mathrm{a}_\ell) = \emptyset$ *and* $\mathrm{dchild}(\mathrm{a}_\ell) = \emptyset$.

We refer to the root of the tree as the 0th level of the tree and count levels in the standard way. The height of the tree is the number of levels in the tree. Note that a diffusion tree defined in this way can readily be interpreted as a branching process (See, e.g., Chapter 0 in [14]), in which the jth generation of the process corresponds with the jth level nodes in Tr.

Next we incorporate the good behavior variable G_t with the diffusion tree. The motivation is that, roughly speaking, consistently good behavior guarantees a small number of infections, or creation of children, at each level. This can be seen through Lemma 2.2.

DEFINITION 2.6. (STOPPED DIFFUSION TREE)
Consider a diffusion process with island diffusion rule, and let $T(\ell)$ *be the time that* a_ℓ *becomes infected in the process. The stopped diffusion tree* Tr' *(with respect to* a_i *and* $\mathcal{W}$*) is a subtree of* Tr *induced by the set of vertices* $\{\mathrm{a}_\ell : \mathrm{a}_\ell \in \mathrm{Tr} \ \wedge \ G_{T(\ell)} = 1\}$. *We write* $\mathrm{a}_\ell \in \mathrm{child}'(\mathrm{a}_{\ell'})$ *if* $\mathrm{a}_\ell \in \mathrm{child}(\mathrm{a}_{\ell'})$ *and* $\mathrm{a}_\ell \in \mathrm{Tr}'$. *Also,* $\mathrm{a}_\ell \in \mathrm{dchild}'(\mathrm{a}_{\ell'})$ *if* $\mathrm{a}_\ell \in \mathrm{dchild}(\mathrm{a}_{\ell'})$ *and* $\mathrm{a}_\ell \in \mathrm{Tr}'$.

Note that the definition of the stopped diffusion tree involves global behavior of the whole diffusion process due to the introduction of G_t. On the other hand, $\mathrm{Tr} = \mathrm{Tr}'$ with overwhelming probability, so we can translate the properties of Tr' back to Tr easily. The main property of the stopped diffusion tree is the following:

LEMMA 2.3. *Consider a diffusion process with the island diffusion rule. Let* a_ℓ *be an arbitrary agent with infection time* $T(\ell)$. *We have*

$$\mathrm{E}\left[|\mathrm{dchild}'(\mathrm{a}_\ell)| \middle| \mathcal{F}_{T(\ell)}, S^\ell_{T(\ell)} \in \mathfrak{V}(2\ell_2 \log^{-4} n)\right] \leq \log^{-3} n, \tag{2.3}$$

where $\mathrm{dchild}'(\cdot)$ *is defined for a stopped diffusion tree with respect to an arbitrary set* $\mathcal{W} \subseteq \mathfrak{V}(20\ell_2 \log n)$.

We regard the conditional expectation in Equation 2.3 as a random variable, and $\mathcal{F}_t$ is the filtration up to time t. The interpretation is that the expected number of a_ℓ's direct children is less than $\log^{-3} n$, regardless of the global configuration at the infection time of a_ℓ, as long as it lies in $\mathfrak{V}(2\ell_2 \log^{-4} n)$ at that time.

Recursive utilization of Lemma 2.3 on successive tree levels leads to the following lemma:

LEMMA 2.4. *Consider a diffusion process with the island diffusion rule starting with an infected island* $\mathrm{Isd}_0(\mathrm{a}_1, \ell_1 \log^{-1} n)$. *For the stopped diffusion tree* Tr' *with respect to any* $\mathcal{W} \subseteq \mathfrak{V}(20\ell_2 \log n)$, *let* $\mathrm{Height}(\mathrm{Tr}')$ *be its height. Then we have* $\Pr[\mathrm{Height}(\mathrm{Tr}') > 2\log n] \leq \exp(-3 \log n \log\log n)$.

This lemma is the key to proving Proposition 2.1. The bound on the height of the diffusion tree limits the number of infected generations from the initially infected a_1. Together with short travel distance property in Definition 2.3, it effectively constrains the positions of all infected agents into a small ball around a_1, which leads to Proposition 2.1.

2.2 From local to global process This subsection will be devoted to proving Theorem 2.1 via Proposition 2.1, or in other words, to turn our local probabilistic bound into a global result on the diffusion time.

We note that Proposition 2.1 deals with the case when there is only one initially infected agent. As discussed briefly in the discussion following the proposition, we want to iterate this estimate so that at every time increment Δt, the infected region is constrained within a certain radius from the initial positions of all the agents that are already infected at the start of the increment. Our argument is aided by noting which agents infect other agents. To ease the notation for this purpose, we introduce an artificial concept of virus type, denoted by $\nu_{i,t}$. We say an agent gets a virus of type $\nu_{i,t}$ if the meeting events of this agent can be traced upstream to the agent a_i, where a_i is already infected at time t. In other words, assume that a_i is infected at time t, and imagine that we remove the viruses in all infected agents except a_i but we keep the same dynamics of all the random walks. We say a particular agent gets $\nu_{i,t}$ if it eventually gets infected under this imaginary scenario. Note that under this artificial framework of virus types it is obvious that an agent can get many different types of virus, in terms of both i and t.

In parallel to Proposition 2.1, we introduce the family of binary random variables $b_{i,t}$ to represent whether a virus of type $\nu_{i,t}$ can be constrained in a ball with radius $9\ell_2 \log n$:

DEFINITION 2.7. ($b_{i,t}$ AND VIRUS OF TYPE $\nu_{i,t}$) *Let* $\overline{\mathfrak{B}} = \mathbb{B}(P, \frac{n}{4})$ *where* $P = (n/2, n/2, n/2)$. *Let* $\mathrm{a}_1, ..., \mathrm{a}_m$ *be agents that are uniformly distributed on* $\mathcal{V}^3$ *at* $t = 0$ *and diffuse according to Definition 1.1. Let* t *be an arbitrary time step and* $i \in [m]$. *At time* t, *a virus of type* $\nu_{i,t}$ *emerges on agent* a_i *and diffuses. Define the binary random variable* $b_{i,t}$ *as follows:*

- If $S^i_t \in \overline{\mathfrak{B}}$: $b_{i,t}$ *is set as* 1 *if and only if all the agents infected by the virus of type* $\nu_{i,t}$ *at time* $t + \Delta t$ *can be covered by the ball* $\mathbb{B}(S^i_t, 9\ell_2 \log n)$.
- If $S^i_t \notin \overline{\mathfrak{B}}$: $b_{i,t} = 1$.

Let us start with the observation that $b_{i,t} = 1$ for all i and t with high probability. This can be seen easily by applying Proposition 2.1 to every agent and taking a union bound across all i and t.

COROLLARY 2.1. *Consider the family of random variables* $\{b_{i,t} : i \in [m], t \leq n^{2.5}\}$ *defined above. We have*

$$\Pr\left[\bigwedge_{i\in[m], t\leq n^{2.5}} (b_{i,t} = 1)\right] \geq 1 - \exp(-4 \log n \log\log n).$$

We also need the following lemma on the densities of agents in a linearly sized ball centered at $(n/2, n/2, n/2)$, which can be shown using a standard Chernoff bound and taking a union bound across all t.

LEMMA 2.5. *Let* $\mathfrak{B} = \mathbb{B}(P, n/8)$, *where* $P = (n/2, n/2, n/2)$. *Let* B_t *be the indicator variable that there is at least one agent in* $\mathfrak{B}$ *at time* t. *Let* $B = \prod_{t \le n^{2.5}} B_t$, *the indicator variable that there is at least one agent in* $\mathfrak{B}$ *at all times in* $[0, n^{2.5}]$. *We have*

$$\Pr[B = 0] \le \exp(-\log^2 n)$$

for sufficiently large n.

We next present our major lemma for this subsection.

LEMMA 2.6. *Let* $\mathrm{a}_1, ..., \mathrm{a}_m$ *be placed uniformly at random on* $\mathcal{V}^3$ *such that* $m \ge 1600 n \log^2 n$. *Let* $\ell_2 = \sqrt{n^3/m}$. *Let* $\{b_{i,t} : i \in [m], t \le n^{2.5}\}$ *and* B *be the random variables described above. If* $b_{i,t} = 1$ *for all* i, t *and* $B = 1$, *then the diffusion time is at least* $T_c = \frac{1}{81}\ell_2 n \log^{-29} n$.

Notice that by Corollary 2.1 and Lemma 2.5,

$$\Pr\left[\bigwedge_{i \le m, t \le n^{2.5}} (b_{i,t} = 1)\right] \ge 1 - \exp(-4 \log n \log\log n).$$
$$\Pr[B = 1] \ge 1 - \exp(-\log^2 n).$$

Together with Lemma 2.6, Theorem 2.1 then follows.

Proof. Without loss of generality, we assume the x, y, and z coordinates of S_0^1 are all negative. We can always rotate the space $\mathcal{V}^3$ at $t = 0$ correspondingly to ensure this assumption holds.

We shall prove by contradiction. Consider two balls $\mathfrak{B}$ and $\overline{\mathfrak{B}}$ defined above. Assume the diffusion time is less than T_c. First, because $B = 1$, a necessary condition for the diffusion to complete is that an infected agent visits the smaller ball $\mathfrak{B}$ at a time $T' \le T_c$ (since otherwise the agents in $\mathfrak{B}$ would be uninfected all the time, including at T_c). We call this agent $\mathrm{a}_{i'}$. Next, for the infection to get into $\mathfrak{B}$, it must happen that there is an infected agent that enters $\overline{\mathfrak{B}}$ from outside, whose infection trajectory eventually reaches $\mathrm{a}_{i'}$. We denote T'' to be the *last* time that this happens, and the responsible agent to be $\mathrm{a}_{i''}$. We focus on the trajectory of infection that goes from $\mathrm{a}_{i''}$ to $\mathrm{a}_{i'}$ that lies completely inside $\overline{\mathfrak{B}}$ (which exists since T'' is the last time of entry). Note that we consider at most $\lceil T_c/\Delta t \rceil$ time increments of Δt. Now, since $b_{i,t} = 1$ for all i and t, by repeated use of triangle inequality, we get

$$\begin{aligned}
\|S_{T'}^{i'} - S_{T''}^{i''}\|_\infty &\le 9\ell_2 \log n \left\lceil \frac{T_c}{\Delta t} \right\rceil \\
&\le 9\ell_2 \log n \left(\frac{(1/81)\ell_2 n \log^{-29} n}{\ell_2^2 \log^{-28} n} + 1 \right) \\
&\le \frac{n}{9} + 9\ell_2 \log n \\
&< \frac{n}{8} - 1
\end{aligned}$$

On the other hand, the physical dimensions of $\mathfrak{B}$ and $\overline{\mathfrak{B}}$ give that

$$\|S_{T'}^{i'} - S_{T''}^{i''}\|_\infty \ge \frac{n}{8} - 1$$

which gives a contradiction.

3 Upper bound

We now focus on an upper bound for the diffusion time. Our main result is the following:

THEOREM 3.1. *Let* $a_1, \ldots, a_m$ *be placed uniformly at random on* $\mathcal{V}^3$, *where* $n \le m \le n^3$. *Let* $\hat{\ell}_2 = \sqrt{n^3/m} \log n$. *When* n *is sufficiently large, the diffusion time* T *satisfies*

$$\Pr[T \ge 128 n \hat{\ell}_2 \log^{47} n] \le \exp(-\frac{1}{2}\log^2 n).$$

Note that this theorem shows that an upper bound of $\tilde{O}(n\sqrt{n^3/m})$ holds for the diffusion time with high probability. Hence the upper and lower bounds "match" up to logarithmic factors. We remark that the constant 47 in the exponent has not been optimized.

The main goal of this section is to prove this theorem. Our proof strategy relies on calculating the growth rate of the *total* infected agents evolving over time; such growth rate turns out to be best characterized as the increase/decrease in infected/uninfected agents relative to the size of the corresponding population. More precisely, we show that for a well-chosen time increment, either the number of infected agents doubles or the number of uninfected agents reduces by half with high probability. The choice of time increment is complex, depending on the analysis of the local interactions in small cubes and the global geometric arrangements of these cubes with respect to the distribution of infected agents.

As with the lower bound proof, our technique for proving Theorem 3.1 is different from existing methods. Roughly, existing methods can be decomposed into two steps (see for example [10]): 1) In the first step, consider a small ball of length r that contains the initially infected agent. One can see that for $d = 2$, when the number of agents in the ball is $\tilde{\Theta}(m(r/n)^2)$, within time increment r^2 the number of infections to agents initially in this ball is $\tilde{\Omega}(1)$ w.h.p.. 2) The second step is to prove that for any ball that has $\tilde{\Omega}(1)$ infected agents at time t, its surrounding adjacent balls will also have $\tilde{\Omega}(1)$ infected agents by time $t + r^2$. From these two steps, one can recursively estimate the time to spread infection across the whole space $\mathcal{V}^2$ to be $n/r \times r^2 = nr$ w.h.p.. In other words, at time nr all the balls in $\mathcal{V}^2$ will have $\tilde{\Omega}(1)$ infected agents. Moreover, every agent in

$\mathcal{V}^2$ is infected in the same order of time units, because $\tilde{\Omega}(1)$ is also the total number of agents in any ball under good density condition. Finally, it is then clear that a good choice of r is then $n/\sqrt{m}$, which would give the optimal upper bound.

The critical difference in the analysis for $d > 2$ lies primarily in the magnitude of the meeting probability of random walks. In the case of $d = 2$, the meeting probability of two random walks at distance r within time r^2 is $\tilde{\Theta}(1)$, whereas for $d > 2$ the meeting probability is $\Theta(1/r^{d-2})$. For $d = 2$, this means that it is easy, i.e. w.h.p., for infection to transmit from a ball with $\tilde{\Omega}(1)$ infected agents to an adjacent uninfected ball, so that the latter also has $\tilde{\Omega}(1)$ infected agents after a time increment of r^2. In the case $d > 2$, however, $\Omega(r^{d-2})$ infected agents must be present in a ball to transmit virus effectively to its adjacent uninfected ball within r^2 time. Consequently, arguing for transmission across adjacent balls becomes problematic (more details are in the full paper). In light of this, we take an alternate approach to analyze both the local interactions and the global distribution of infected agents. Instead of focusing on transmission from one infected ball to another, we calculate the spreading rate across the whole space. This turns out to be fruitful in obtaining a tight upper bound.

We briefly describe the forthcoming analysis. As with the lower bound, we start with local analysis. We partition the space $\mathcal{V}^3$ into disjoint subcubes each of size $\hat{\ell}_2 \times \hat{\ell}_2 \times \hat{\ell}_2$. Here $\hat{\ell}_2$ is just a logarithmic factor larger than ℓ_2, the size of subcubes used for the lower bound, so that with overwhelming probability there are *at least* $\hat{\ell}_2$ agents in a subcube. We show that, within every subcube, over a time increment of length $\Theta(\hat{\ell}_2^2)$ the number of infections is roughly a $\tilde{\Omega}(1)$-fraction of the minimum of the number of infected and uninfected agents. Hence, at least locally, we have the desired behavior described above.

We then leverage the local analysis to obtain the global result. However, this is not straightforward. For example, consider the beginning when the number of infected agents is small. If infected agents are distributed uniformly throughout the whole space, it would be easy to show that new infections would roughly grow in proportion to the number of infected agents. However, if infected agents are concentrated into a small number of subcubes, we have to show that there are enough neighboring subcubes on the boundary of these infected subcubes that these subcubes become infected suitably rapidly, so that after the appropriate time increment the number of infected agents doubles. Similar arguments arise for the case when infected agents are dominant, with the end result being a halving of the uninfected population.

We now make the above discussion rigorous. First, let $b = (2n+1)/\hat{\ell}_2$, so there are in total b^3 subcubes. As in the previous section, we divide the time into small intervals. We reuse the symbol Δt to represent the length of each interval but here we set $\Delta t = 16\hat{\ell}_2^2$. Our local bound is built within each subcube (and pair of neighboring subcubes) in the time increment Δt:

LEMMA 3.1. *Let $\mathcal{W} \subset \mathcal{V}^3$ be a region that can be covered by a ball of radius $2\hat{\ell}_2$ under the L_∞-norm. Let A^f and A^u be subsets of infected and uninfected agents in $\mathcal{W}$ at time t such that $|A^f| = m_1$, $|A^u| = m_2$, and $\max\{m_1, m_2\} = \hat{\ell}_2/\log^2 n$. Given any initial placement of the agents of A^f and A^u, let $M(t)$ be the number of agents in A^u that become infected at time $t + \Delta t$. We have $\Pr\left[M(t) \geq \frac{\tau_0 \min\{m_1,m_2\}}{\log^4 n}\middle|\mathcal{F}_t\right] \geq \tau_0 \log^{-6} n$ for some constant τ_0, where $\mathcal{F}_t$ denotes the information of the whole diffusion process up to time t.*

The high level idea in proving Lemma 3.1 is to count the total number of times the infected agents meet the uninfected ones between time t and $t + \Delta t$. The probability two agents in $\mathcal{W}$ can meet each other within time Δt is approximately $\tilde{\Omega}(1/\hat{\ell}_2)$ (this fact is proved in the full paper). The expected number of meetings is thus $\tilde{\Omega}(1/\hat{\ell}_2) \times m_1 m_2 = \tilde{\Omega}(\min\{m_1, m_2\})$. The total number of newly infected agents is the number of meetings modulo possible overcounts on each originally uninfected agent. Hence we show that the number of meetings is $\tilde{O}(1)$ for each uninfected agent to conclude that $\tilde{\Omega}(\min\{m_1, m_2\})$ more agents become infected at time $t + \Delta t$.

The next step is to characterize the growth rate at a larger scale. This requires more notation. We denote the set of b^3 subcubes of size $\hat{\ell}_2 \times \hat{\ell}_2 \times \hat{\ell}_2$ as $\mathfrak{C} = \{h_{i,j,k} : i, j, k \in [b]\}$. For an arbitrary subcube $h_{i,j,k}$, we define its neighbors as $N(h_{i,j,k}) = \{h_{i',j',k'} : |i - i'| + |j - j'| + |k - k'| = 1\}$. In other words, $h_{i',j',k'}$ is a neighbor of $h_{i,j,k}$ if and only if both subcubes share a facet. Let $\mathcal{H}$ be an arbitrary subset of $\mathfrak{C}$. We write $N(\mathcal{H}) = \bigcup_{h\in\mathcal{H}} N(h)$.

DEFINITION 3.1. (EXTERIOR AND INTERIOR SURFACE) *Let $\mathcal{H}$ be a subset of $\mathfrak{C}$. The* exterior surface *of $\mathcal{H}$ is $\partial\mathcal{H} = N(\mathcal{H}) - \mathcal{H}$. Let $\overline{\mathcal{H}}$ be the complement of $\mathcal{H}$. The* interior surface *of $\mathcal{H}$ is $\partial\overline{\mathcal{H}} = N(\overline{\mathcal{H}}) - \overline{\mathcal{H}}$, i.e., the exterior surface of the complement of $\mathcal{H}$.*

At time step $t = i\Delta t$, let $\mathcal{G}_t$ be the set of all subcubes that contain more than $\hat{\ell}_2/2$ infected agents and let $g_t = |\mathcal{G}_t|$; let $\mathcal{B}_t = \overline{\mathcal{G}_t}$ be the rest of the subcubes and let $b_t = |\mathcal{B}_t|$. We call a subcube in $\mathcal{G}_t$ an *infected*

(*good*) subcube and a subcube in $\mathcal{B}_t$ an *uninfected* (*bad*) subcube.

We classify the agents in the process according to the subcubes they reside in. To facilitate our analysis, we adopt the notational system $\mathfrak{A}_t$ and $\mathfrak{A}_t^{\cdot,\cdot}$ to represent the total number of agents that belong to the type specified in the superscript. Specifically, let $\mathfrak{A}_t^f$ be the set of infected agents at time t; decompose the set $\mathfrak{A}_t^f$ as $\mathfrak{A}_t^f = \mathfrak{A}_t^{f,\mathcal{G}} \cup \mathfrak{A}_t^{f,\mathcal{B}}$, where $\mathfrak{A}_t^{f,\mathcal{G}}$ is the set of infected agents residing in the subcubes in $\mathcal{G}_t$ and $\mathfrak{A}_t^{f,\mathcal{B}}$ the set of infected agents in $\mathcal{B}_t$. Similarly, let $\mathfrak{A}_t^u$ be the set of all uninfected agents; decompose the set $\mathfrak{A}_t^u$ as $\mathfrak{A}_t^u = \mathfrak{A}_t^{u,\mathcal{G}} \cup \mathfrak{A}_t^{u,\mathcal{B}}$, where $\mathfrak{A}_t^{u,\mathcal{G}}$ is the set of uninfected agents residing in the subcubes in $\mathcal{G}_t$ and $\mathfrak{A}_t^{u,\mathcal{B}}$ the set of uninfected agents in $\mathcal{B}_t$. Furthermore, we denote $\Delta\mathfrak{A}_t^{\mathcal{G}}$ and $\Delta\mathfrak{A}_t^{\mathcal{B}}$ as the set of agents in $\mathcal{G}_t$ and $\mathcal{B}_t$ respectively that are infected between t and $t + \Delta t$. Hence the total increase in infected agents, or equivalently the total decrease in uninfected agents, between t and $t + \Delta t$ is given by $\Delta\mathfrak{A}_t = \Delta\mathfrak{A}_t^{\mathcal{G}} \cup \Delta\mathfrak{A}_t^{\mathcal{B}}$. Lastly, we let $\widetilde{\Delta\mathfrak{A}_t^{\mathcal{G}}}$ be the set of agents in $\mathcal{G}_t \cup \partial\mathcal{G}_t$ that are infected between t and $t + \Delta t$.

Similar to the lower bound analysis, here we also introduce good density conditions that can be easily verified to hold with high probability, and reuse the symbols D_t and D with slightly different meanings from the last section:

DEFINITION 3.2. *Let $\{D_t : t \geq 0\}$ be a sequence of binary random variables such that $D_t = 1$ if for all time steps on or before t, the number of agents for any subcube in $\mathcal{V}^3$ with size $\hat{\ell}_2 \times \hat{\ell}_2 \times \hat{\ell}_2$ is between $\hat{\ell}_2$ and $2\hat{\ell}_2 \log^2 n$. Also, let $D = D_{n^{2.5}}$.*

The following lemma shows that $D_t = 1$ with high probability:

LEMMA 3.2. *For any $t \leq n^{2.5}$, $\Pr[D_t = 0] \leq \exp(-\frac{1}{15}\log^2 n)$ for sufficiently large n.*

We now state two bounds on the growth rate of the agent types, one relative to the "boundary subcubes" $\partial\mathcal{G}_t$ and one relative to the total agents of each type:

COROLLARY 3.1. *For some constant τ_0,*

$$\Pr\left[|\widetilde{\Delta\mathfrak{A}_t^{\mathcal{G}}} \cap \Delta\mathfrak{A}_t^{\mathcal{B}}| \geq |\partial\mathcal{G}_t| \cdot \frac{\tau_0\hat{\ell}_2}{4\log^{13} n}\middle|\mathcal{F}_t, D_t = 1\right] \geq \tau_0 \log^{-6} n.$$

Consequently,

$$\Pr\left[|\widetilde{\Delta\mathfrak{A}_t^{\mathcal{G}}}| \geq |\partial\mathcal{G}_t| \cdot \frac{\tau_0\hat{\ell}_2}{4\log^{13} n}\middle|\mathcal{F}_t, D_t = 1\right] \geq \tau_0 \log^{-6} n$$

and

$$\Pr\left[|\Delta\mathfrak{A}_t^{\mathcal{B}}| \geq |\partial\mathcal{G}_t| \cdot \frac{\tau_0\hat{\ell}_2}{4\log^{13} n}\middle|\mathcal{F}_t, D_t = 1\right] \geq \tau_0 \log^{-6} n.$$

COROLLARY 3.2. *We have*

$$\Pr\left[|\Delta\mathfrak{A}_t^{\mathcal{G}}| \geq \frac{\tau_0^2}{4\log^{38} n}|\mathfrak{A}_t^{u,\mathcal{G}}|\middle|\mathcal{F}_t, D_t = 1\right] \geq \tau_0 \log^{-6} n$$

and

$$\Pr\left[|\Delta\mathfrak{A}_t^{\mathcal{B}}| \geq \frac{\tau_0^2}{4\log^{38} n}|\mathfrak{A}_t^{f,\mathcal{B}}|\middle|\mathcal{F}_t, D_t = 1\right] \geq \tau_0 \log^{-6} n.$$

The proofs of these two corollaries both rely on using coupled diffusion processes that have slower diffusion rates. These processes only allow infection locally i.e. within each "pair" of subcubes on the surface of $\mathcal{G}_t$ in the case of Corollary 3.1 and within each subcube in Corollary 3.2, and hence can be tackled by Lemma 3.1. The surface $\partial\mathcal{G}_t$ in Corollary 3.1 appears naturally from a matching argument between neighboring infected and uninfected subcubes. Roughly speaking, the bounds in Corollary 3.1 are tighter and hence more useful for the cases where infected/uninfected agents are dense in the infected/uninfected subcubes, while those in Corollary 3.2 are for cases where the agent types are more uniformly distributed.

We now move to the global diffusion upper bound. As discussed in the beginning of this section, the balance between the distributions of each type of subcube and the distributions of actual agents plays a crucial role in our analysis. Fix an arbitrary time t, we classify the joint configurations of the agents into four types:

- type 1 (namely $\mathcal{P}_{1,t}$): when $|\mathcal{G}_t| \leq \frac{1}{2}((2n+1)/\hat{\ell}_2)^3$ and $|\mathfrak{A}_t^{f,\mathcal{G}}| \geq \frac{1}{2}|\mathfrak{A}_t^f|$.
- type 2 (namely $\mathcal{P}_{2,t}$): when $|\mathcal{G}_t| \leq \frac{1}{2}((2n+1)/\hat{\ell}_2)^3$ and $|\mathfrak{A}_t^{f,\mathcal{G}}| < \frac{1}{2}|\mathfrak{A}_t^f|$.
- type 3 (namely $\mathcal{P}_{3,t}$): when $|\mathcal{G}_t| > \frac{1}{2}((2n+1)/\hat{\ell}_2)^3$ and $|\mathfrak{A}_t^{u,\mathcal{G}}| < \frac{1}{2}|\mathfrak{A}_t^u|$.
- type 4 (namely $\mathcal{P}_{4,t}$): when $|\mathcal{G}_t| > \frac{1}{2}((2n+1)/\hat{\ell}_2)^3$ and $|\mathfrak{A}_t^{u,\mathcal{G}}| \geq \frac{1}{2}|\mathfrak{A}_t^u|$.

Recall that $\mathcal{F}_t$ refers to the information on the global configurations up to time t. We shall abuse notation slightly and say $\mathcal{F}_t \in \mathcal{P}_{i,t}$ if the configuration of the agents at time t belongs to the ith type described above. Notice that $\mathcal{F}_t$ belongs to exactly one of the sets $\mathcal{P}_{1,t}$, $\mathcal{P}_{2,t}$, $\mathcal{P}_{3,t}, \mathcal{P}_{4,t}$. In brief, scenarios $\mathcal{P}_{1,t}$ and $\mathcal{P}_{2,t}$ have a majority of uninfected subcubes, while $\mathcal{P}_{3,t}$ and $\mathcal{P}_{4,t}$ have a majority of infected subcubes. From another perspective, $\mathcal{P}_{1,t}$ and $\mathcal{P}_{3,t}$ refer to situations when the dominant types (with respect to the status of infection) are dense in their subcube types (infected/uninfected

subcubes), while $\mathcal{P}_{2,t}$ and $\mathcal{P}_{4,t}$ refer to the more uniform scenarios. The next lemma states that when $\mathcal{F}_t \in \mathcal{P}_{1,t} \cup \mathcal{P}_{2,t}$, the total number of infected agents $|\mathfrak{A}_t^f|$ grows in proportion to a monotone function of $|\mathfrak{A}_t^f|$ within Δt steps. On the other hand, when $\mathcal{F}_t \in \mathcal{P}_{3,t} \cup \mathcal{P}_{4,t}$, the total number of uninfected agents $|\mathfrak{A}_t^u|$ is reduced in proportion to a monotone function of $|\mathfrak{A}_t^u|$ within Δt steps.

LEMMA 3.3. *Fix an arbitrary t, define the following events,*

$$\begin{aligned}
e_1(t) &= \left\{ |\Delta\mathfrak{A}_t| \geq 0.09\tau_0 \left(\frac{|\mathfrak{A}_t^f|}{4\hat{\ell}_2 \log^2 n} \right)^{2/3} \frac{\hat{\ell}_2}{\log^{13} n} \right\} \\
e_2(t) &= \left\{ |\Delta\mathfrak{A}_t| \geq \frac{\tau_0^2}{8\log^{38} n} |\mathfrak{A}_t^f| \right\} \\
e_3(t) &= \left\{ |\Delta\mathfrak{A}_t| \geq 0.015\tau_0 \left(\frac{|\mathfrak{A}_t^u|}{4\hat{\ell}_2 \log^2 n} \right)^{2/3} \frac{\hat{\ell}_2}{\log^{13} n} \right\} \\
e_4(t) &= \left\{ |\Delta\mathfrak{A}_t| \geq \frac{\tau_0^2}{8\log^{38} n} |\mathfrak{A}_t^u| \right\}.
\end{aligned}$$

We have $\Pr[e_i \mid \mathcal{F}_t \in \mathcal{P}_{i,t}, D_t = 1] \geq \tau_0 \log^{-6} n$ *for* $i = 1, 2, 3, 4$.

Intuitively, e_1 and e_2 connect the number of newly infected agents to the original number of infected agents. When e_1 or e_2 are triggered sufficiently many times, the number of infected agents doubles. Meanwhile, e_3 and e_4 connect the number of newly infected agents to the original number of uninfected agents. When e_3 or e_4 are triggered sufficiently many times, the number of uninfected agents halves.

The key to proving Lemma 3.3, which will ultimately lead to a bound on the global growth rate of doubling/halving the total number of infected/uninfected agents as depicted in the next proposition, is a geometric relation between the boundary of $\mathcal{G}_t$, i.e. $\partial\mathcal{G}_t$, and $\mathcal{G}_t$ itself. More specifically, an isoperimetric bound on $\mathcal{G}_t$ guarantees that no matter how packed together these good subcubes are, there are still an order $|\mathcal{G}_t|^{2/3}$ of them exposed to the bad subcubes, hence the global infection rate cannot be too slow.

Our major proposition presented next essentially pins down the number of times these events need to be triggered to double the number of infected agents or halve the number of uninfected ones.

PROPOSITION 3.1. *Consider the information diffusion problem over $\mathcal{V}^3$ with m agents. For any fixed $t \leq n^{2.5} - 4\sqrt{\frac{m}{n}} \log^{45} n\Delta t$, define the following events*

$$\chi_1(t) \equiv \left(|\mathfrak{A}^f_{t+4\sqrt{\frac{m}{n}} \log^{45} n\Delta t}| \geq 2|\mathfrak{A}_t^f| \right) \text{ and}$$

$$\chi_2(t) \equiv \left(|\mathfrak{A}^u_{t+4\sqrt{\frac{m}{n}} \log^{45} n\Delta t}| \leq \frac{1}{2}|\mathfrak{A}_t^u| \right).$$

We have $\Pr[\chi_1(t) \vee \chi_2(t)] \geq 1 - \exp(-\log^2 n)$.

Note that this bound suggests that for each time increment $4\sqrt{\frac{m}{n}} \log^{45} n\Delta t$, either the number of infected agents doubles or the number of uninfected agents is reduced by half with high probability. Therefore, within time at most $2\log n \cdot \left(4\sqrt{\frac{m}{n}} \log^{45} n\Delta t\right) = 128n\hat{\ell}_2 \log^{47} n$ all the agents get infected with probability at least $1 - 2\log n \exp(-\log^2 n)$. This proves Theorem 3.1.

To summarize our approach, Corollaries 3.1 and 3.2 first translate the local infection rate of Lemma 3.1 into a rate based on the subcube types (i.e. good and bad subcubes). Then Lemma 3.3 further aggregates the growth rate to depend only on the infected and uninfected agents, by looking at the geometrical arrangement of the subcubes. Nevertheless, the bound from Lemma 3.3 is still too crude, but by making a long enough sequence of trials i.e. $4\sqrt{\frac{m}{n}} \log^{45} n$ times, at least one of the four scenarios defined in Lemma 3.3 occurs for a significant number of times, despite the $\Omega(\log^{-6} n)$ probability of occurrence for each individual step for any of the four scenarios. This leads to the probabilistic bound for $\chi_1(t) \vee \chi_2(t)$.

4 The case when the number of agents is sparse

Finally, let us state our result on the case where $m = o(n)$:

PROPOSITION 4.1. *Let $\mathrm{a}_1, \mathrm{a}_2, ..., \mathrm{a}_m$ be placed uniformly at random on $\mathcal{V}^3$, where $m < n\log^{-2} n$. Let a_1 be the agent that holds a virus at $t = 0$, and T be the diffusion time. We have for any constant $c > 0$,*

$$\Pr[T < \frac{n^3}{m} \log^{-c} n] \leq \log^{-c} n \text{ and}$$

$$\Pr[T > \frac{2n^3}{m} \log^{15} n] \leq \exp(-(\log^2 n)/2).$$

References

[1] O. S. M. Alves, F. P. Machado, S. Yu. Popov, and K. Ravishankar. The shape theorem for the frog model with random initial configuration. *arXiv:math/0110280v1*, October 2001.

[2] O. S. M. Alves, S. Yu. Popov, and F. P. Machado. The shape theorem for the frog model. *Annals of Applied Probability*, 12(2):533–546, 2002.

[3] Tamio Arai, Eiichi Yoshida, and Jun Ota. Information diffusion by local communication of multiple mobile robots. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics, 1993*, pages 535–540, 1993.

[4] Andrea E. F. Clementi, Angelo Monti, and Riccardo Silvestri. Fast flooding over manhattan. In *PODC*, pages 375–383. ACM, 2010.

[5] Andrea E. F. Clementi, Francesco Pasquale, and Riccardo Silvestri. Manets: High mobility can make up for low transmission power. In *ICALP (2)*, volume 5556 of *Lecture Notes in Computer Science*, pages 387–398. Springer, 2009.

[6] Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97:1090–1098, 2001.

[7] Harry Kesten and Vladas Sidoravicius. The spread of a rumor or infection in a moving population. *Annals of Probability*, 33(6):pp. 2402–2462, 2005.

[8] Abdelmajid Khelil, Christian Becker, Jing Tian, and Kurt Rothermel. An epidemic model for information diffusion in manets. In *Proceedings of the 5th ACM international workshop on Modeling analysis and simulation of wireless and mobile systems*, MSWiM, pages 54–60. ACM, 2002.

[9] Yuval Peres, Alistair Sinclair, Perla Sousi, and Alexandre Stauffer:. Mobile geometric graphs: Detection, coverage and percolation. In *SODA*, pages 412–428. SIAM, 2011.

[10] Alberto Pettarin, Andrea Pietracaprina, Geppino Pucci, and Eli Upfal. Tight bounds on information dissemination in sparse mobile networks. *arXiv:1101.4609v2*, 2011.

[11] Purnamrita Sarkar and Andrew W. Moore. Dynamic social network analysis using latent space models. *SIGKDD Explorations*, 7(2):31–40, 2005.

[12] Alistair Sinclair and Alexandre Stauffer. Mobile geometric graphs, and detection and communication problems in mobile wireless networks. *arXiv:1005.1117v2*, 2010.

[13] Yi Wang, Shyam Kapadia, and Bhaskar Krishnamachari. Infection spread in wireless networks with random and adversarial node mobilities. In *Proceeding of the 1st ACM SIGMOBILE workshop on Mobility models*, MobilityModels '08, pages 17–24. ACM, 2008.

[14] David Williams. *Probability with martingales*. Cambridge mathematical textbooks. Cambridge University Press, 1991.