Epidemic Spreading with External Agents

Siddhartha Banerjee, Aditya Gopalan, Abhik Kumar Das, and Sanjay Shakkottai

***Abstract*—We study the spread of epidemics in large networks when the spread is assisted by a small number of *external agents*: infection sources with bounded spreading power, but whose movement is unrestricted vis-à-vis the underlying network topology. For networks which are *'spatially constrained'*, we show that the spread of infection can be significantly speeded up even by a few such agents infecting *randomly*. More specifically, for general networks with bounded external virulence (e.g., a single or finite number of random mobile agents), we derive upper-bounds on the order of the infection time as a function of network size. Conversely, for certain common classes of networks such as line graphs, grids and random geometric graphs, we also derive lower bounds on the order of the spreading time over all (potentially network-state aware and adversarial) external infection-spreading policies; these adversarial lower bounds match (up to logarithmic factors) the spreading time achieved by an external agent with random mobility. This demonstrates that random, state-oblivious infection-spreading by an external agent is in fact order-wise optimal for dissemination in such spatially constrained networks.**

***Index Terms*—Epidemic processes, infection/information spreading, long-range contact, mobility.**

I. Introduction

Various natural and engineered phenomena around us involve the spread of information or infection through different kinds of networks. Rumours and news stories propagate among people linked by various means of communication, diseases diffuse as epidemics through populations by various modes, plants disperse pollen/seeds and thus genetic traits geographically, riots spread across pockets of communities, advertisers aim to disseminate information about goods through networks of consumers, and computer viruses, email worms and software patches piggyback across computer networks. Understanding how infection/information/innovation can travel across networks through such *epidemic processes* has been a subject of extensive study in disciplines ranging from epidemiology [2], [3], [4], sociology [5], [6] and computer science [7], [8], [9] to physics [10], information theory/networking [11], [12], [13], [14], [15] and applied mathematics [16], [17], [18], yielding valuable insights into qualitative and quantitative aspects of spreading behaviour in networks.

At a high level, the spread of such infections in networks is due to the interaction of two processes-(i) a local spreading process in the underlying network, and (ii) a global infection process due to agents that are external to the network. Based on this observation, we propose and study a model for spreading in networks that decomposes into two distinct components – a basic *intrinsic spread* component in which infection spreads *locally* among neighbouring nodes due to the underlying graph topology, and an additional *external spread* component in which 'external agents' (potentially unconstrained by the underlying graph) can carry infection far from its origin, helping it spread *globally*. More specifically, we develop a rigorous framework with which we quantify the effect that a number of *omniscient* (*i.e.* network-state aware) and *adversarial* (*i.e.* attempting to maximize the rate of infection) external infection agents can have on the time required to spread infection throughout the network.

We stress that the generic terms 'intrinsic spread' and 'external spread' serve to model a variety of situations involving heterogeneous modes of spreading. In the context of wireless communication, for instance, consider the increasingly studied propagation [19], [20], [21], [22] of viruses and worms that exploit the connectivity afforded by both modern short-range personal communication technologies like Bluetooth and long-range media such as SMS/MMS and the Internet. To paraphrase Kleinberg [23], outbreaks due to such worms are well-modelled by local spreading on a fixed network of nodes in space (*i.e.* short-range Bluetooth wireless transmissions between neighbouring quasi-static users) aided by relatively unrestricted paths through the network (*i.e.* long-range, faster-timescale emails and messages through SMS/MMS/Internet). Other, more classically-studied, examples of multiscale spreading include those of natural disease epidemics [2], [24] and bioterror attacks [25], where infection can spread locally through spatial pathways (*i.e.* interpersonal contact) and through large-scale geographic means (*e.g.* human movement through airline routes) [26]. In all these and allied cases, a form of agency, external to the underlying graph, is responsible for long-range proliferation of an otherwise locally diffusive contagion, and it is the effect of this external agency that we wish to investigate.

Given the applicability of our epidemic model with external-infection sources, a fundamental characterization of the impact of external agency on epidemic spread has a twofold utility:

(a) *(Adversarial perspective)* Whenever malicious epidemics (such as those described before) threaten to spread via both intrinsic and external means, it becomes important to understand the *worst-case* long-range spreading behaviour (this is the component that can potentially accelerate the spread) in order to deploy appropriate countermeasures.

(b) *(Optimization perspective)* In cases where propagation is desirable and the external component can be *controlled*, an adversarial study of external-agent assisted spreading

An earlier version of this work appears in the proceedings of IEEE Infocom, Shanghai, China, April 2011 [1]. This work was partially supported by NSF Grants CNS-1017525, CNS-0721380 and Army Research Office Grant W911NF-11-1-0265.

S. Banerjee, A. K. Das and S. Shakkottai are with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX-78705, USA (e-mail: siddhartha@utexas.edu, akdas@utexas.edu, shakkott@austin.utexas.edu).

A. Gopalan is with the Faculty of Electrical Engineering at the Technion-Israel Institute of Technology, Haifa, Israel (e-mail: aditya@ee.technion.ac.il).

can constructively help design fast mobile spread strategies (*e.g.* viral advertising [9], network protocol design [27] and diffusion of innovations [5]).

A. Main Contributions

We consider large graphs $G = (V, E)$ in which an epidemic starts at a designated node and commences spreading through two interacting processes: an *intrinsic spread* which follows the standard Susceptible-Infected (SI) dynamics [11] (also termed *contact process* [16], [17]) with *i.i.d.* exponentially distributed propagation times for each edge, and an additional *external infection*. To model the external infection process, we allow *every* node in the graph to get infected at a potentially different (including zero) exponential rate at each instant; thus at time t, the state of the network consists of a set of nodes which are infected (therefore determining the intrinsic infection process) and a $|V|$-dimensional vector $\bar{L}(t)$ of external infection-rates for each node. The *virulence* of the external agents is measured by $||\bar{L}(t)||_1$ (i.e., the sum rate of external infection). We allow $\bar{L}(t)$ to be chosen as a function of the network state and history (*omniscience*) and further, chosen *adversarially*, i.e., designed to minimize the time taken to infect all nodes; in Section II we discuss how this model generalizes various models for long-range infection spreading.

In this setting, our main message is somewhat surprising– in spite of the 'adversarial power' external agents have for spreading infection, we show that for common *spatially-constrained* graphs (i.e., having high diameter/low conductance), a simple random strategy is order-optimal. More formally, our contributions in this paper are as follows:

(a) We develop general *upper bounds* on the order of the infection time (expectation and concentration results) for large graphs when the external infection pattern is *random*, i.e. when every node is susceptible to the same external-infection rate, irrespective of the infection-state and graph topology. The bounds are based on the graph topology (in particular, diameter/conductance of appropriate subgraphs) and a lower bound $L_{\min}(n)$ on the virulence (which can scale with the network size). We also analyze an alternate *greedy* infection policy based on the same graph partitioning scheme, for which we obtain better bounds.

(b) For common classes of structured graphs (ring/line graphs, d-dimensional grids) and random graphs (geometric random graphs) which have high diameter/low conductance (spatially-constrained), and given an upper bound $L_{\max}(n)$ on the virulence, we use first-passage percolation theory [17] to derive *lower bounds* on the order of infection times (again, both in expectation and w.h.p) over *all* (possibly omniscient and adversarial) external-infection policies. These lower bounds match the upper bounds for random spreading up to logarithmic factors, showing that *random external-infection policies are order-wise optimal* for such spatially-constrained graphs. Furthermore, they match exactly for the greedy policies, indicating that these bounds are tight.

(c) Apart from results for specific graphs and policies, the general bounds (and related techniques) are of independent interest in that they provide a fairly complete picture of the dependence of spreading time on external virulence and graph topology in a wide regime; in particular, it is tight for graphs with *polynomially-bounded diameter* (i.e., diameter $D(n) = \Omega(n^\alpha)$ for some $\alpha > 0$) and sub-linear external virulence (i.e.,$||L(t)||_1 = o(n)$). To demonstrate this, we discuss how other spreading models (graphs with additional static or dynamic edges, and/or mobile agents) can be analyzed in our framework, and what our bounds translate to in such cases.

B. Related Work

Prior work concerning network spread, though diverse in scope and treatment, does not address the impact of adversarial external-agent-assisted spreading in spatially-constrained networks. Moreover, to the best of our knowledge, it lacks a consistent analytical framework in which the effects of different forms of external agency on spreading time can be compared. There has been much work in studying the static spread of infection/innovation using various notions of influence and susceptibility, both numerically using field data/extensive simulations [10], [5], [6], [7], [28] and analytically [11], [16], [17], [18], [29]. For the case of spreading via external agents, many numerical studies have investigated the spread of infectious diseases with specific mobility patterns, *e.g.* via airline networks [30], [26], heterogeneous geographic means [2], [3], and recently, electronic pathways [23], [19], [20], [21], [22]. Several notable works in communication engineering include studies in which all network nodes are simultaneously mobile– for designing gossip algorithms [15], [27] or improving the capacity of wireless networks [14] –and analysis of rumour spreading on fully-connected graphs [12], [13]. Other design-oriented studies include the investigation of optimal seeding in networks for maximum spread from a computational perspective [9], efficient routing over spatial networks with fixed long-range links [8] and tight bounds on deterministic spreading with external-agents in d-dimensional hypercubes [31].

II. Model for Epidemics with External Agents

Consider a sequence of graphs $G_n = (V_n, E_n)$ indexed by n, with the n-th graph having n nodes (for ease of notation, we label the nodes in V from 1 to n). For instance, G_n could be the ring graph with n nodes, or a (2-dimensional) $\sqrt{n} \times \sqrt{n}$ grid. For convenience, we will drop the subscript n for all quantities pertaining to the graph G_n when the context is clear.

We model the spread of an epidemic on the underlying graph G_n (or G) using a continuous-time *spreading process* $(S(t))_{t\geq 0}$. At each time t, $S(t) = (S_1(t), \ldots, S_n(t)) \in \{0,1\}^V$ denotes the 'infection state' of the nodes in V: $S_i(t) = 0$ (resp. $S_i(t) = 1$) indicates that node $i \in V$ is 'healthy' (resp. 'infected') at time t. Let us denote by $\mathcal{S}(t)$ the set of infected nodes at time t, *i.e.* $\mathcal{S}(t) \stackrel{\triangle}{=} \{i \in V : S_i(t) = 1\}$, and use $\mathcal{N}(S(t))$ to denote its size. In order to capture the

effect of external agents, the evolution of $S(t)$ is assumed to be driven by the following modes of infection spread:

- *Intrinsic infection:* Initially, at $t = 0$, all nodes are healthy, except for a single node (node 1, say) that is infected. Once any node is infected, it always remains infected, and infects each of its neighbouring (w.r.t. G) healthy nodes at random times, independent and exponentially distributed with mean 1. We term this form of spread as infection via *basic* contact.
- *External infection:* At time t, in addition to being infected by its neighbours in G, each node i is susceptible to an *external infection* (or alternately, a *long-range contact*) with an exponential infection-rate $L_i(t)$. The external infection-rate vector $\bar{L}(t) \equiv (L_i(t))_{i \in V}$ can vary with time t and can depend on the state of the network $S(t)$

We note here that the dependence of the external infection rate $\bar{L}(t)$ on the network state allows us to model the propagation of infection through a wide range of external infection processes transcending the structure of the underlying network (G). For instance,

(a) $\bar{L}(t) = 0$ represents infection occurring only through edges of the underlying graph (the standard *contact process*).

(b) Well-known studies by Kleinberg [32] and Watts-Strogatz [33] show that adding a few fixed *long-range* edges onto structured networks can dramatically reduce routing time and diameter. The above long-range contact model captures the dynamics of infection spreading with L such additional edges, say, by letting $L_i(t)$ be the number of long-range edges incident on node i that have an infected node at the other end at time t.

(c) Long-range edges over the underlying graph, instead of being drawn in a *static* manner, can be *dynamically* added and deleted as time progresses. For instance, infected nodes can "throw out" fresh sets of long-range edges at certain times –this corresponds to choosing fresh sets of long-range infection targets depending on network state or other parameters.

(d) Moving beyond long-range structures, the external infection vector can also be used to model "virtual mobility"; the external infection could be caused by one or several mobile agents, whose position is unconstrained by the graph, and which thus spread infection to various parts of the network with corresponding rates $\bar{L}(t)$.

(e) At an even more abstract level, the external agent can be viewed as an external information source with *bandwidth* $||\bar{L}(t)||_1$, which can share its bandwidth across nodes of the graph. Such a model can be used to design optimal spreading processes for viral advertising, spread of software updates, etc.

Throughout this paper, we have used the term *intrinsic infection* (likewise *external infection*) interchangeably with *basic contact* (likewise *long-range contact*). Now, to complete our system description, we term the quantity $||\bar{L}(t)||_1$ as the long-range *virulence* at time t, i.e., the power of the infection to spread through long-range contact. In this work, we restrict ourselves to scenarios where $L(t)$ is uniformly upper and lower bounded by functions $L_{\max}(n), L_{\min}(n)$ respectively (that can potentially scale with the network size n). Finally, we define the *finish time* or *spreading time* of the epidemic as $T \stackrel{\triangle}{=} \inf\{t \geq 0 : S(t) = \mathbb{1}_n\}$, i.e., the time at which all nodes in V get infected. Our concerns are both to (a) analyze the finish time under certain natural long-range spreading dynamics, and (b) show universal lower-bounds on the finish time for common structured networks, over a wide class of long-range spreading dynamics.

General Notation: We use $\mathbb{Z}$ and $\mathbb{R}$ for the set of integers and reals respectively, and also use standard Landau notation (O, Θ, Ω) for the asymptotic growth rate of functions. For random variables X and Y, the notation $X \leq_{st} Y$ and $Y \geq_{st} X$ means that Y stochastically dominates X, *i.e.* $\mathbb{P}[Y \geq z] \geq \mathbb{P}[X \geq z]$ for all z. Where necessary, we follow the convention that $1/\infty \stackrel{\triangle}{=} 0$.

III. Main Results and Discussion

We now state our main results, and discuss what they translate to for different models of local epidemics aided by spreading via external agents.

We state our results for general external virulence $L(t)$ (more specifically, in terms of the bounds $L_{\min}(n)$ and $L_{\max}(n)$ on $||L(t)||_1$) in two parts: upper bounds for spreading time for *general graphs* under specific policies (in particular, a random policy and a greedy policy), and lower bounds under *any policy* for certain *specific graphs* (in particular, rings/line graphs, d-dimensional grids and the geometric random graph), which are spatially-constrained, and where the bounds are tight. We conclude the section with a discussion of the applicability of our bounds and techniques; in particular, we show how our bounds can be used to obtain results on spreading time for various models of external infection such as long-range links and mobile agents, and discuss their limitations.

A. Upper Bounds for Specific Policies

Our first main result is an upper-bound on the finish time (both in expectation and with high probability) of the homogeneous external virulence policy, or as we refer to it hereafter, the *random spread policy*, for a general graph G. Such a policy is equivalent to one in which the (single) external agent chooses a node uniformly at random and starts infecting it; hence the name 'random spreading policy'. The following result states that if G can be broken into a (large) number of uniformly-sized pieces, then the time taken by random spreading to finish is of the order of the number of pieces or the piece size, whichever dominates.

Theorem 1 (Upper bound for Homogeneous External Virulence: Diameter version)**.** *Suppose $L_i(t) = L_j(t)$, $i, j = 1, \ldots, n$, and $||\bar{L}(t)||_1 \geq L_{\min}(n) \geq 0$ for all $t \geq 0$. Suppose also that for each n, the graph G_n admits a partition $G_n = \bigcup_{i=1}^{g(n)} G_{n,i}$ by $g(n)$ connected subgraphs $G_{n,i}$, each with size $\Theta(s(n))$ and diameter $O(d(n))$. Then,*

(a) *(Mean finish time) $\mathbb{E}[T] = O(h(n) \log n)$, where $h(n) \equiv \max\left(\frac{g(n)}{L_{\min}(n)}, d(n)\right)$.*

(b) *(Finish time concentration) If $g(n) = \Omega(n^{\delta})$ for some $\delta > 0$, then for any $\gamma > 0$ there exists $\kappa = \kappa(\gamma) > 0$ such that $\mathbb{P}[T \geq \kappa h(n) \log n] = O(n^{-\gamma})$.*

To understand how this result is applied, consider a line graph on n nodes–this can be partitioned into $\sqrt{n}$ segments of length (diameter) $\sqrt{n}$ each, and hence by the above result, the random spreading policy takes $O(\sqrt{n} \log n)$ time to infect all nodes. We formally state and derive such results subsequently.

Next we obtain a spreading time bound for a *greedy* spreading policy, which we call the *Greedy Subgraph Infection* (or GSI) policy. The policy is based on the (optimal) partitioning of the graph that we constructed in the above theorems, and is as follows: given the subgraphs $G_i, i \in \{1, 2, \ldots, g(n)\}$, they are infected through sequential greedy (as opposed to *homogeneous*) long-range contact, i.e., $||L(t)||_1 = L_{\min}(n)$, and $L(t)$ is supported on a single node $j(t)$ within any *maximally healthy* subgraph at time t (i.e., one which minimizes $|G_i \cap \mathcal{S}(t)|$). The finish time of the GSI policy is $O(h(n))$ in expectation and *w.h.p.*, which we state as follows:

Theorem 2 (Upper bound for GSI Policy)**.** *Suppose for each n, the graph G_n admits partition $\bigcup_{i=1}^{g(n)} G_{n,i}$ of connected subgraphs $G_{n,i}$, each of size $\Theta(s(n))$ and diameter $O(d(n))$. Further, $d(n) = \log n + \omega(1)$. Then for spreading via the Greedy Subgraph Infection policy, we have $\mathbb{E}[T] = O(h(n))$, where $h(n) \equiv \max\left(\frac{g(n)}{L_{\min}(n)}, d(n)\right)$.*

Again, applying this to the line graph with n nodes, we now get a spreading time of $O(\sqrt{n})$, which improves on the previous bound by a factor of $\log n$.

Our final upper bound is an alternate bound for the finish time with random external-agents in terms of a different structural property intimately related to spreading ability in graphs – the *conductance* (also called the *isoperimetric constant*). The conductance $\Psi(G)$ of a graph $G = (V, E)$ is defined as

$$\Psi(G) \stackrel{\triangle}{=} \inf_{S \subset V : 1 \leq S \leq \frac{|V|}{2}} \frac{E(S, V \setminus S)}{|S|},$$

where for $A, B \subseteq V$, $E(A, B)$ denotes the number of edges that have exactly one endpoint each in A and B. The conductance of a graph is a widely studied measure of how fast a random walk on the graph converges to stationarity [34], [29].Analogous to Theorem 1, the next result formalizes the idea that spreading on a graph is dominated by the larger of (a) the number of pieces it can be broken into, and (b) the reciprocal of the piece conductance.

Theorem 3 (Upper Bound for Homogeneous External Virulence: Conductance version)**.** *Suppose that for each n, the graph G_n admits a partition $G_n = \bigcup_{i=1}^{g(n)} G_{n,i}$ by $g(n)$ connected subgraphs $G_{n,i}$, each with size $\Theta(s(n))$ and conductance $\Theta(\Psi(n))$. Then,*

(a) *(Mean finish time) $\mathbb{E}[T_{\pi_r}] = O(k(n) \log g(n))$, where $k(n) \equiv \max\left(\frac{g(n)}{L_{\min}(n)}, \frac{\log s(n)}{\Psi(n)}\right)$.*

(b) *(Finish time concentration) There exists $\kappa > 0$ independent of n such that*

$$\mathbb{P}[T_{\pi_r} \geq \kappa k(n) \log g(n)] = O\left((\log g(n))^{-2}\right).$$

B. Results: Lower Bounds for Specific Topologies

Having estimated the spreading time of random and greedy external-infection policies, a natural question that arises at this point is: How do these policies compare with the best (possibly omniscient and adversarial) policy, *i.e.*, with the lowest possible spreading time among *all other infection strategies*? To this end, we show that for certain commonly studied *spatially-limited* networks (i.e., with diameter $\Omega(n^{\alpha})$ for some $\alpha > 0$), such as line/ring networks, d-dimensional grids and random geometric graphs, random spreading yields the best order-wise time up to a logarithmic factor (and the GSI policy yields the best order-wise time) to spread infection. In particular, for each of these classes of graphs, we establish lower bounds on the finish time of *any* spreading strategy, that match the upper bounds established in the previous section.

Rings/Linear Graphs: Let $G_n = (V_n, E_n)$ be the ring graph with n contiguous vertices $V_n \stackrel{\triangle}{=} \{v_1, \ldots, v_n\}$, $E_n \stackrel{\triangle}{=} \{(v_i, v_j) : j - i \equiv 1 \pmod{n}\}$. By partitioning G_n into $\sqrt{n}$[1] successive $\sqrt{n}$-sized segments, where the diameter of each segment is $\sqrt{n}$, an application of Theorem 1 gives:

Corollary 1 (Time for random spread on ring graphs)**.** *For the random spread policy π_r on the ring/line graph G_n, we have*

(a) $\mathbb{E}[T_{\pi_r}] = O\left(\sqrt{\frac{n}{L_{\min}(n)}} \log n\right)$,

(b) *For any $\gamma > 0$ $\exists \alpha = \alpha(\gamma) > 0$ such that $\mathbb{P}\left[T_{\pi_r} \geq \alpha \sqrt{\frac{n}{L_{\min}(n)}} \log n\right] = O(n^{-\gamma})$.*

i.e., the finish time on an n-ring, with random mobility, is $O(\sqrt{n} \log n)$ in expectation and with high probability.

Our next main result is the following lower bound for the spreading time of an epidemic on ring graphs due to any external-infection policy.

Theorem 4 (Lower bound for ring graphs)**.** *For the ring graph G_n with n nodes, and given that $||\bar{L}(t)||_1 \leq 1 \, \forall t \geq 0$, there exists $c > 0$ independent of n such that for any spreading policy π,*

$$\mathbb{P}\left[T_{\pi} < c\sqrt{n}\right] = O\left(e^{-\Theta(1)\sqrt{n}}\right).$$

Moreover, we have

$$\inf_{\pi \in \Pi} \mathbb{E}[T_{\pi}] = \Omega\left(\sqrt{n}\right).$$

d-dimensional Grids: Building on the previous result, we next show that the random spread strategy achieves the order-wise optimal finish time even on d-dimensional grid networks where $d \geq 2$. Given d, the d-dimensional grid graph $G_n = (V_n, E_n)$ on n nodes is given by $V_n \stackrel{\triangle}{=} \{1, 2, \ldots, n^{1/d}\}^d$, and $E_n \stackrel{\triangle}{=} \{(x, y) \in V_n \times V_n : ||x - y||_1 = 1\}$.

Consider a partition of G_n into $(n/L_{\min})^{1/(d+1)}$ identical and contiguous 'sub-grids' $G_{n,i}$, $i = 1, \ldots, n^{1/(d+1)}$ (for details, refer to Section V-B). With such a partition, an application of Theorem 1 shows that

[1]For ease of notation, we assume that fractional powers of n take integer value; if not, the bounds can be modified by appropriately taking ceiling/floor.

Corollary 2 (Time for random spread on d-dimensional grids)**.** *For the random spread policy π_r on an n-node d-dimensional grid G_n, we have*

(a) $\mathbb{E}[T_{\pi_r}] = O\Big(\left(\frac{n}{L_{\min}(n)}\right)^{1/(d+1)} \log n\Big)$,

(b) *For any $\gamma > 0$ there exists $\alpha = \alpha(\gamma) > 0$ with* $\mathbb{P}\Big[T_{\pi_r} \geq \alpha \left(\frac{n}{L_{\min}(n)}\right)^{1/(d+1)} \log n\Big] = O(n^{-\gamma})$.

i.e., the finish time with random external-infection on a d-dimensional n-node grid is $O\left(\left(\frac{n}{L_{\min}(n)}\right)^{1/(d+1)} \log n\right)$ in expectation and with high probability.

Further we show that *any* external-infection policy on a grid must take time $\Omega\left(\left(\frac{n}{L_{\max}(n)}\right)^{1/(d+1)}\right)$ to finish infecting all nodes with high probability, and consequently also in expectation, thereby showing the above bound is order-optimal.

Theorem 5 (Lower bound: d-dimensional grids, bounded long-range virulence)**.** *Let G_n be a symmetric n-node d-dimensional grid graph. Suppose that $||\bar{L}(t)||_1 \leq L_{\max}(n) = \omega(n)$ for all $t \geq 0$. Then, there exist $c_1, c_2 > 0$, not depending on n, such that*

$$\mathbb{P}\left[T \leq c_1 \left(\frac{n}{L_{\max}(n)}\right)^{\frac{1}{d+1}}\right] = O\left(e^{-c_2\left(\frac{n}{L_{\max}(n)}\right)^{\frac{1}{2d+2}}}\right).$$

Furthermore, if $L_{\max}(n) = O(n^{1-\epsilon})$ for some $\epsilon \in (0, 1]$, then

$$\mathbb{E}[T] = \Omega\left(\left(\frac{n}{L_{\max}(n)}\right)^{\frac{1}{d+1}}\right).$$

Geometric Random Graphs: Finally, we shift focus from structured graphs to a popular family of random graphs, widely used for modelling physical networks. The *Geometric Random Graph (RGG)* is a random graph model wherein n points (*i.e.* nodes) are placed i.i.d. uniformly in $[0, 1] \times [0, 1]$. Two nodes x, y are connected by an edge iff $||x - y|| \leq r_n$, where r_n is often called the *coverage radius*. The RGG $G_n = G_n(r_n)$ consists of the n nodes and edges as above.

It is known that when the coverage radius r_n is above a critical threshold of $\sqrt{\log n/\pi}$, the RGG is connected with high probability [35]. In our last set of results, we show that similar to before, random spreading on RGGs in this critical connectivity regime is optimal upto logarithmic factors. First, we show with high probability that random spreading finishes in time $O(\sqrt[3]{n} \log n)$:

Theorem 6 (Time for random spread on RGGs)**.** *For the planar random geometric graph $G_n(r_n)$, if $r_n \geq \sqrt{\frac{5 \log n}{n}}$, then there exists $\alpha > 0$ such that*

$$\lim_{n\to\infty} \mathbb{P}[T_{\pi_r} \geq \alpha \sqrt[3]{n/L_{\min}(n)} \log n] = 0.$$

Finally, we follow this up with a converse result that shows that no other policy can better this time (order-wise, up to the logarithmic factor) with significant probability. This directly parallels the earlier results about finish times on 2-dimensional grids, where random mobile spread exhibits the same optimal order of growth.

Theorem 7 (Lower bound for RGGs)**.** *For the planar geometric random graph G_n with $r_n = O(\sqrt{\log n/n})$ and any spreading policy π with $L_{\max}(n) = O(n^{1-\epsilon})$ for some $\epsilon \in (0, 1]$, $\exists\ \beta > 0$ such that $\lim_{n\to\infty} \mathbb{P}\left[T_\pi \geq \beta \frac{\sqrt[3]{n/L_{\max}(n)}}{\log^{4/3} n}\right] = 1$.*

C. Discussion and Extensions

The framework of epidemic spreading with external agents encompasses many known models for epidemic spreading with long-range contacts (as we discussed previously in Section II): this is done by appropriately specifying $\bar{L}(t) \in \mathbb{R}_+^{|V|}$ as a function of time t, network topology and network-state $S(t)$. For example, the presence of a single additional 'static long-range' link $(i, j) \in V^2$ is equivalent to setting $L_i(t) = \beta \mathbb{1}_{S_j(t)=1}, L_j(t) = \beta \mathbb{1}_{S_i(t)=1}$ and $L_k(t) = 0 \forall k \notin \{i, j\}$ (where β is the rate of spreading on the edge). We now discuss the implications of our results and techniques on such models of long-range contact.

Static Links: To demonstrate our results in the context of a graph overladen with additional static edges, consider a d-dimensional grid with $L(n)$ additional static links. Then we have the following lower-bound for the spreading time T (obtained by setting $L_{\max}(n) = L(n)$ in Theorem 5).

Corollary 3. *Let G_n be a symmetric n-node d-dimensional grid graph, with $L(n)$ additional static links. If $L(n) = O(n^{1-\epsilon})$ for some $\epsilon \in (0, 1]$, then $\mathbb{E}[T] = \Omega\left(\left(\frac{n}{L(n)}\right)^{\frac{1}{d+1}}\right)$.*

Note that by combining this with Theorem 2, we can also get the same lower bound on the diameter $D(n)$ of the resultant graph. To see this, observe that by considering the entire graph as a single partition, Theorem 2 gives that the spreading time is $O(D(n))$, and thus $D(n) = \Omega\left((n/L(n))^{\frac{1}{d+1}}\right)$ by the above Corollary. One consequence of this is in the context of 'small-world graphs' [32], [33] wherein the diameter of a d-dimensional grid on n nodes is reduced to $\Theta(\log n)$ by adding $\Omega(n \log n)$ *random* long-range edges. The usefulness of the above result is to show that it is not possible to obtain such sub-polynomial diameters by adding $O(n^{1-\epsilon})$ edges.

We note also that this bound is tight. We can see this from the following simple example: partition the graph into $L(n)$ identical segments, and add an edge between a chosen vertex i and a single vertex in each segment. Now for an epidemic starting at node i, it is easy to see that the resultant process is equivalent to the 2-phase spreading process in the proof of Theorem 2 (i.e., parallel seeding of clusters followed by local spreading in clusters). Hence, the spreading time for this process is $O\left((n/L(n))^{\frac{1}{d+1}}\right)$.

Dynamic Links and Mobile Agents: A more surprising result is obtained by considering spreading on a grid with additional *dynamic* links, i.e., long-range links which can change their endpoints as time progresses. Unlike a static link which can transmit the infection only once (before both its endpoints are infected), such dynamic links can be re-used over time to help spread the infection. However, we can again use Theorem 5 to get the following

Corollary 4. *Let G_n be a symmetric n-node d-dimensional grid graph, with $L = O(n^{1-\epsilon}), \epsilon \in (0,1]$ additional dynamic links. Then $\mathbb{E}[T] = \Omega\left(\left(\frac{n}{L}\right)^{\frac{1}{d+1}}\right)$.*

Thus, using dynamic links does not reduce the order of the spreading time.

Related to dynamic links are models of epidemic spreading via *mobile agents*–in such a context, assuming $L(n)$ mobile agents, each with constant infection-rate, Theorem 5 again gives the same converse for spreading time, i.e., $\Omega\left((n/L(n))^{\frac{1}{d+1}}\right)$ for d-dimensional grids. Furthermore, the techniques of Theorems 1 and 2 can be used to give upper bounds for various models of mobility: for example, for L mobiles moving randomly on a d-dimensional grid (where each mobile is unconstrained by the graph as to its next location), Theorem 1 shows that the spreading time is $O\left((n/L(n))^{\frac{1}{d+1}}\right)$.

IV. Spreading-Time Bounds for Specific Policies

In this section, we formally prove the upper bounds on spreading time we stated in Section III-A. We first prove the diameter-based upper-bound on the finish time of the random spread policy for a general graph G.

Theorem (Theorem 1: Upper bound for Homogeneous External Virulence: Diameter version)**.** *Suppose $L_i(t) = L_j(t)$, $i, j = 1, \ldots, n$, and $||\bar{L}(t)||_1 \geq L_{\min}(n) \geq 0$ for all $t \geq 0$. Suppose also that for each n, the graph G_n admits a partition $G_n = \bigcup_{i=1}^{g(n)} G_{n,i}$ by $g(n)$ connected subgraphs $G_{n,i}$, each with size $\Theta(s(n))$ and diameter $O(d(n))$. Then,*

(a) *(Mean finish time) $\mathbb{E}[T] = O(h(n)\log n)$, where $h(n) \equiv \max\left(\frac{g(n)}{L_{\min}(n)}, d(n)\right)$.*

(b) *(Finish time concentration) If $g(n) = \Omega(n^{\delta})$ for some $\delta > 0$, then for any $\gamma > 0$ there exists $\kappa = \kappa(\gamma) > 0$ such that $\mathbb{P}[T \geq \kappa h(n)\log n] = O(n^{-\gamma})$.*

In other words, given any partition of a large graph, the spreading time of an epidemic process, assisted by random external infection, is determined by both (a) the time taken for spread to start in each segment of the partition and (b) the worst possible time taken by the intrinsic spread within each segment.

Proof: The proof uses techniques from (a) stochastic majorization and (b) graph-partitioning into suitable *shortest-path* spanning trees.

Under the homogeneous (or random) external infection policy, we have that $L_i(t) \geq L_{\min}(n)/n$ for all $i = 1, \ldots, n$. As before, $(S(t))_{t\geq 0}$ denotes the infection state process. Now observe that each subgraph $G_{n,i} \equiv G_i$ is prone to infection (i.e., some node in $G_{n,i}$ contracts infection) due to long-range contact with an exponential rate of $\Omega\left(L_{\min}(n) \cdot s(n)/n\right)$. Consider an alternative infection-spreading process $(\tilde{S}(t))_{t\geq 0}$ which evolves in two phases:

- *Phase-1:* Throughout this phase, infection spread occurs only due to long-range contact, and not through basic/static contact. The phase starts from time $t = 0$ and ends when at least one node in each subgraph G_i is infected. Let T_1 be the end time of this phase.
- *Phase-2:* In this phase, infection spread occurs only due to basic contact through neighbours in G, and not through long-rang contact. At time T_1, for each subgraph G_i, only the *first node* infected in phase-1, say N_i, is assumed to be infected, and all other nodes in G_i are considered healthy, even if some of them were infected on phase-1 subsequent to N_i. The process $\tilde{S}(\cdot)$ proceeds from time T_1 onwards by the usual static spread dynamics *within* each G_i, *i.e.* with the restriction that infection *does not spread* across edges that connect different subgraphs. Denote by T_2 the additional time taken (since T_1) for all nodes in all the G_i to get infected.

A standard coupling argument establishes that $\mathcal{N}(S(t))$ stochastically dominates $\mathcal{N}(\tilde{S}(t))$ for all t, i.e., $\tilde{S}$ is a 'slower' process than S. Thus, the finish time for $\tilde{S}(\cdot)$ stochastically dominates that of $S(\cdot)$, *i.e.*

$$T \leq_{st} T_1 + T_2. \tag{1}$$

It remains to estimate the means of T_1 and T_2, and their tail probabilities, to finish the proof. The analysis for T_1 follows a standard coupon-collecting argument: memorylessness of the exponential distribution implies that T_1 is stochastically dominated by the maximum of $g(n)$ *i.i.d.* exponential random variables with parameter $\Omega(L_{\min}(n) \cdot s(n)/n)$, i.e., $\Omega(L_{\min}(n)/g(n))$. Hence, using a well-known result about the expectation of the maximum of *i.i.d.* exponentials, we obtain

$$\begin{aligned}\mathbb{E}[T_1] &= O\left(\frac{H_{g(n)}}{L_{\min}(n)/g(n)}\right)\\ &= O(g(n)\log g(n)/L_{\min}(n)),\end{aligned} \tag{2}$$

where $H_k \stackrel{\triangle}{=} \sum_{i=1}^{k} i^{-1} = O(\log k)$ is the kth harmonic number. Also, by a union bound over the tails of $g(n)$ *i.i.d.* exponential random variables, for any $\kappa > 0$ we can estimate the tail of T_1:

$$\begin{aligned}&\mathbb{P}[T_1 \geq \kappa g(n)\log g(n)]\\ &\qquad \leq g(n) e^{-(\Theta(L_{\min}(n)/g(n))\kappa g(n)\log g(n))}\\ &\qquad = g(n)^{-\Theta(\kappa L_{\min}(n))+1}.\end{aligned} \tag{3}$$

To estimate the statistics of T_2, we further consider the following 'slower' mode of (static) spread than that of phase-2: for each subgraph G_i (with diameter $O(d(n))$), let W_i be a *shortest-path spanning tree* of G_i rooted at the node N_i which is infected in phase-1. Such a tree has diameter $O(d(n))$ and can, in principle, be obtained by performing a Breadth-First Search (BFS) on G_i starting at N_i. If we now insist that the phase-2 static infection process in G_i spreads *only via the edges of* W_i, then again, a standard coupling can be used to show that the time $\tilde{T}_2$ when all nodes in G get infected thus stochastically dominates T_2.

Before proceeding, we need the following simple lemma, which we state without proof:

Lemma 1. *For real numbers a_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n$, $\max_{i=1}^{m}\sum_{j=1}^{n} a_{ij} \leq \sum_{j=1}^{n}\max_{i=1}^{m} a_{ij}$.*

Now for each tree W_i, let its leaves be labelled $N_{i1}, \ldots, N_{il(i)}$. Each leaf N_{ij} has a unique path p_{ij} starting

from N_i to itself, of length $O(d(n))$. Let $\hat{T}_{jk}$ be the time taken for the infection to spread across the kth edge on this path p_{ij}, *i.e.* the (exponentially distributed) interval between the times when the $(k-1)$th node and the kth node on the path are infected. Then, the time $\hat{T}_{2,i}$ taken for all nodes in W_i (hence G_i) to get infected can be upper-bounded by using Lemma 1:

$$\hat{T}_{2,i} = \max_{j=1}^{l(i)} \sum_{k=1}^{|p_{ij}|} \hat{T}_{jk} \leq \sum_{k=1}^{O(d(n))} \left(\max_{j=1}^{l(i)} \hat{T}_{jk} \right),$$

and a further application of the lemma bounds the phase-2 finish time $\hat{T}_2 = \max_{i=1}^{g(n)} \hat{T}_{2,i}$ as

$$\begin{aligned} \hat{T}_2 &\leq \max_{i=1}^{g(n)} \sum_{k=1}^{O(d(n))} \left(\max_{j=1}^{l(i)} \hat{T}_{jk} \right) \\ &\leq \sum_{k=1}^{O(d(n))} \left(\max_{i=1}^{g(n)} \max_{j=1}^{l(i)} \hat{T}_{jk} \right). \end{aligned}$$

The term in brackets is simply the maximum of the infection spread times across all stage-k edges of all the trees W_i within G. Hence, it is stochastically bounded above by the maximum of n i.i.d Exponential(1) random variables (say $Z_1, \ldots, Z_n$), using which we can write

$$\mathbb{E}[T_2] \leq \mathbb{E}[\hat{T}_2] \leq \sum_{k=1}^{O(d(n))} O(H_n) = O(d(n) \log n). \quad (4)$$

Again, using the union bound to estimate the tail probability of T_2, we have, for any $\kappa > 0$,

$$\begin{aligned} \mathbb{P}[T_2 \geq \kappa d(n) \log n] &\leq \mathbb{P}[\hat{T}_2 \geq \kappa d(n) \log n] \\ &\leq O(d(n)) \mathbb{P}[Z_1 \geq \kappa \log n] \\ &\leq n \cdot n e^{-\kappa \log n} = n^{-\kappa+2}. \end{aligned} \quad (5)$$

We now have all the required pieces. Combining (1), (2) and (4) with the fact that $g(n) = O(n)$ proves the first part of the theorem. For the second part, the hypothesis that $g(n) = \Omega(n^\delta)$, together with (3), gives

$$\begin{aligned} \mathbb{P}[T_1 \geq \kappa h(n) \log n] &\leq \mathbb{P}[T_1 \geq \kappa g(n) \log g(n)] \\ &\leq n^{-\delta \Theta(\kappa L_{\min}(n)) + \delta}, \end{aligned}$$

which, together with (1) and (5), gives

$$\begin{aligned} \mathbb{P}[T_{\pi_r} \geq 2\kappa h(n) \log n] &\leq \mathbb{P}[T_1 + T_2 \geq 2\kappa h(n) \log n] \\ &\leq n^{-\delta \Theta(\kappa L_{\min}(n)) + \delta} + n^{-\kappa+2} \\ &\leq 2n^{-\min\{\delta(\Theta(\kappa L_{\min}(n)) - 1), \kappa - 2\}} \end{aligned}$$

Choosing κ s.t. $\min\{\delta(\Theta(\kappa L_{\min}(n)) - 1), \kappa - 2\} \geq \gamma$ yields the bound in the second part of the theorem. ∎

The factor of $\log n$ in the bound of the above theorem is actually *only due to* the 'coupon-collector' effect phase-1 time T_1; a more refined analysis of the phase-2 time T_2 shows that if $d(n) = \log n + \omega(1)$, *i.e.* the piece diameter is sufficiently large, then T_2 is order-wise $d(n)$ in expectation and *w.h.p.* This is the intuition behind the spreading time bound for the *Greedy Subgraph Infection* policy. Recall the GSI policy was defined as follows: given the subgraphs G_i, they are infected through sequential greedy (as opposed to *homogeneous*) long-range contact, i.e., $||L(t)||_1 = L_{\min}(n)$, and $L(t)$ is supported on a single node $j(t)$ within any *maximally healthy* subgraph at time t (i.e., one which minimizes $|G_i \cap \mathcal{S}(t)|$). Then we have:

Theorem (Theorem 2: Upper bound for GSI Policy)**.** *Suppose for each n, the graph G_n admits partition $\bigcup_{i=1}^{g(n)} G_{n,i}$ of connected subgraphs $G_{n,i}$, each of size $\Theta(s(n))$ and diameter $O(d(n))$. Further, $d(n) = \log n + \omega(1)$. Then for spreading via the Greedy Subgraph Infection policy, we have $\mathbb{E}[T] = O(h(n))$, where $h(n) \equiv \max\left(\frac{g(n)}{L_{\min}(n)}, d(n)\right)$.*

Proof: Using the same notation as the earlier proof, we consider the slower, two-phase spreading process, such that $T \leq_{st} T_1 + T_2$: in this case however, phase-1 consists of a sequential 'seeding' of each subgraph (it is clear that this is stochastically dominated by the greedy subgraph infection). Thus T_1 now corresponds to the *sum* of $g(n)$ *i.i.d* exponential random variables with parameter $\Omega(L_{\min}(n))$ (i.e., there is no coupon-collector effect), and thus, via standard results, concentrates around around its mean which is $O(g(n)/L_{\min}(n))$. To complete the proof, we need to tighten our previous bound for $\hat{T}_2$ (and hence, T_2), which, using our previous notation, can be written as:

$$\hat{T}_2 = \max_{i=1}^{g(n)} \hat{T}_{2,i} = \max_{i=1}^{g(n)} \max_{j=1}^{l(i)} \sum_{k=1}^{|p_{ij}|} \hat{T}_{jk},$$

i.e., $\hat{T}_2$ is the maximum sum of infection times over all leaves in all trees W_i. Since the total number of leaves in all the trees W_i is at most n, a union bound yields, for any $\alpha > 0$,

$$\mathbb{P}[\hat{T}_2 > \alpha d(n)] \leq n \mathbb{P}\left[\sum_{i=1}^{d(n)} Z_i > \alpha d(n) \right],$$

where all the Z_i are independent Exponential(1) random variables. A Chernoff bounding technique yields

$$\begin{aligned} \mathbb{P}\left[\sum_{i=1}^{d(n)} Z_i > \alpha d(n) \right] &\leq e^{-\psi \alpha d(n)} \mathbb{E}\left[e^{\psi \sum_{i=1}^{d(n)} Z_i} \right] \\ &= e^{-\psi \alpha d(n)} \left(\mathbb{E}\left[e^{\psi Z_1} \right] \right)^{d(n)} \\ &= e^{-\psi \alpha d(n)} (1-\psi)^{-d(n)} \end{aligned}$$

where $0 \leq \psi < 1$. With $\psi = 1/2$ and any $\alpha > 0$, we have,

$$\mathbb{P}[\hat{T}_2 > \alpha d(n)] \leq n \cdot 2^{d(n)} e^{-\frac{\alpha d(n)}{2}}.$$

Finally, for estimating $\mathbb{E}[\hat{T}_2]$ we have,

$$\begin{aligned} \mathbb{E}[\hat{T}_2] &= \int_0^\infty \mathbb{P}[\hat{T}_2 > x] dx \\ &\leq (2 \log 2 + 2) d(n) \\ &\quad + d(n) \int_{2 \log 2 + 2}^\infty \mathbb{P}[\hat{T}_2 > \alpha d(n)] d\alpha. \\ &\leq 3d(n) + 2^n n d(n) \int_{2 \log 2 + 2}^\infty e^{-\frac{\alpha d(n)}{2}} d\alpha \\ &= 3d(n) + 2ne^{-d(n)} = O(d(n)) \end{aligned}$$

Thus we have $\mathbb{E}[T] = O\left(\max\left(\frac{g(n)}{L_{\min}(n)}, d(n)\right)\right)$. ∎

We conclude this section with the conductance-based upper bound on the finish time with random external-agents.

Theorem (Theorem 3: Upper Bound for Homogeneous External Virulence: Conductance version)**.** *Suppose that for each n, the graph G_n admits a partition $G_n = \bigcup_{i=1}^{g(n)} G_{n,i}$ by $g(n)$ connected subgraphs $G_{n,i}$, each with size $\Theta(s(n))$ and conductance $\Theta(\Psi(n))$. Then,*

(a) *(Mean finish time)* $\mathbb{E}[T_{\pi_r}] = O(k(n)\log g(n))$*, where* $k(n) \equiv \max\left(\frac{g(n)}{L_{\min}(n)}, \frac{\log s(n)}{\Psi(n)}\right)$.

(b) *(Finish time concentration) There exists $\kappa > 0$ independent of n such that*

$$\mathbb{P}[T_{\pi_r} \geq \kappa k(n)\log g(n)] = O\left((\log g(n))^{-2}\right).$$

Proof: As in Theorem 1, we study an associated two-phase spreading process $(\tilde{S}(t))_{t\geq 0}$, where the first phase takes time T_1 to infect at least one node in each G_i, and the infection takes a further time T_2 to spread within every (connected) G_i. A coupling argument establishes that $T_{\pi_r} \leq_{st} T_1 + T_2$.

As before, T_1 is distributed as the maximum of $g(n)$ Exponential($\Theta(L_{\min}(n)/g(n))$) random variables, and standard results yield (for $\kappa > 0$):

$$\mathbb{E}[T_1] = O(g(n)\log g(n)/L_{\min}(n)), \tag{6}$$

and for the variance:

$$\begin{aligned}\text{Var}[T_1] &= \frac{1}{\Theta(L_{\min}(n)^2/g(n)^2)}\sum_{i=1}^{g(n)}\frac{1}{i^2} \\ &= \Theta\left(g(n)^2/L_{\min}(n)^2\right).\end{aligned} \tag{7}$$

Next we have that T_2 is the maximum of the times $T_{2,i}$ for infection to spread in each subgraph G_i. We stochastically dominate each $T_{2,i}$ as follows: for each subgraph G_i, consider a continuous time Markov chain $(\hat{Z}_i)_{t\geq 0}$ on the state space $1,\ldots,|V(G_i)|$ with $\hat{Z}_i(0) = 1$ and transitions $j \to j+1$ at rate $j\Psi(n)$ if $1 \leq j \leq |V(G_i)|/2$, and at rate $(|V(G_i)| - j)\Psi(n)$ if $|V(G_i)|/2 < j \leq |V(G_i)| - 1$. Let $\hat{T}_{2,i}$ be the time taken for the Markov chain $\hat{Z}_i$ to hit its final state $|V(G_i)|$; $\hat{T}_{2,i} = \sum_{j=1}^{|V(G_i)|-1}\hat{T}_{2,i,j}$ where $\hat{T}_{2,i,j}$ is the sojourn time of $\hat{Z}_i$ in state j. We claim that $\hat{T}_{2,i}$ stochastically dominates $T_{2,i}$. To see this, note that at any time t, if the number of infected nodes in the phase-2 spreading process in G_i is $1 \leq j \leq |V(G_i)|/2$, then by the definition of conductance, the rate at which a new healthy node in G_i is infected is at least $j\Psi(n)$. Similarly, if the number of infected nodes is $|V(G_i)|/2 < j < |V(G_i)|$ (*i.e.* the number of healthy nodes is $(|V(G_i)| - j)$), then the rate at which a new healthy node is infected is at least $(|V(G_i)|/2 - j)\Psi(n)$. Hence, by standard Markov chain coupling arguments (e.g., see [11]), we have that $T_{2,i} \leq_{st} \hat{T}_{2,i}$.

By independence of the original phase-2 spreading processes within the G_i for all $i = 1,\ldots,g(n)$, we have

$$\begin{aligned}T_2 = \max_i T_{2,i} \leq_{st} \max_i \hat{T}_{2,i} &= \max_i \sum_{j=1}^{|V(G_i)|-1}\hat{T}_{2,i,j} \\ &\leq \sum_{j=1}^{|V(G_i)|-1}\max_i \hat{T}_{2,i,j}\end{aligned}$$

Hence we have

$$\begin{aligned}\mathbb{E}[T_2] &\leq \sum_{j=1}^{|V(G_i)|-1}\mathbb{E}\left[\max_i \hat{T}_{2,i,j}\right] = 2\sum_{j=1}^{|V(G_i)|/2}\frac{\log g(n)}{j\Psi(n)} \\ &= O\left(\frac{\log s(n)\log g(n)}{\Psi(n)}\right)\end{aligned} \tag{8}$$

Also,

$$\begin{aligned}\text{Var}\left(\sum_{j=1}^{|V(G_i)|-1}\max_i \hat{T}_{2,i,j}\right) &= \sum_{j=1}^{|V(G_i)|-1}\text{Var}\left(\max_i \hat{T}_{2,i,j}\right) \\ = 2\sum_{j=1}^{|V(G_i)|/2}\Theta\left(\frac{1}{j^2\Psi(n)^2}\right) &= \Theta\left(\frac{1}{\Psi(n)^2}\right).\end{aligned} \tag{9}$$

Combining (6) and (8) gives the first part of the theorem. For the second part, we have

$$\begin{aligned}&\mathbb{P}[T_{\pi_r} \geq \kappa k(n)\log g(n)] \leq \mathbb{P}[T_1 + T_2 \geq \kappa k(n)\log g(n)] \\ &\leq \mathbb{P}\left[T_1 + T_2 \geq \frac{\kappa}{2}\left(g(n) + \frac{\log s(n)}{\Psi(n)}\right)\log g(n)\right].\end{aligned}$$

Now, using the variance estimates (7) and (9) with Chebyshev's inequality, we have for large enough $\kappa > 0$

$$\begin{aligned}&\mathbb{P}[T_{\pi_r} \geq \kappa k(n)\log g(n)] \\ &\leq \frac{\text{Var}\left(T_1 + \sum_{j=1}^{|V(G_i)|-1}\max_i \hat{T}_{2,i,j}\right)}{\log^2 g(n)\left(g(n) + \frac{\log s(n)}{\Psi(n)}\right)^2} \\ &= O\left(\frac{\frac{g(n)^2}{L_{\min}(n)^2} + \frac{1}{\Psi(n)^2}}{\log^2 g(n)\left(g(n) + \frac{\log s(n)}{\Psi(n)}\right)^2}\right) \\ &= O\left((\log g(n))^{-2}\right).\end{aligned}$$

This completes the proof. ∎

V. Lower Bounds for Specific Graphs

In the previous section, we have estimated the time taken by random (and greedy) external-infection policies to infect all nodes in a network. In this section, we derive corresponding lower bounds for certain commonly studied *spatially limited* networks, in particular, line/ring networks, d-dimensional grids and random geometric graphs. As discussed in Section III-B, for each of these classes of graphs, we establish lower bounds on the finish time of *any* spreading strategy (possibly omniscient and adversarial) that match the upper bounds (upto logarithmic factors for random spread, and exactly for the GSI policy).

A. Ring/Linear Graphs

As before, for each n we define $G_n = (V_n, E_n)$ to be the ring graph with n contiguous vertices $V_n \triangleq \{v_1, \ldots, v_n\}$, $E_n \triangleq \{(v_i, v_j) : j - i \equiv 1 \pmod n\}$. In the context of Theorem 1, let us partition G_n into $\sqrt{n}$ successive $\sqrt{n}$-sized segments, *i.e.* $G_{n,i}$ is the subgraph induced by $v_{(i-1)\sqrt{n}+1}, \ldots, v_{i\sqrt{n}}$, where i ranges from $1, \ldots, \sqrt{n}$. The diameter of each segment is $\sqrt{n}$, and a straightforward application of the theorem gives

Corollary (Corollary 2:Time for random spread on ring graphs)**.** *For the random spread policy π_r on the ring/line graph G_n, we have*

(a) $\mathbb{E}[T_{\pi_r}] = O\Big(\sqrt{\frac{n}{L_{\min}(n)}}\log n\Big)$,

(b) *For any $\gamma > 0$ $\exists \alpha = \alpha(\gamma) > 0$ such that $\mathbb{P}\Big[T_{\pi_r} \geq \alpha\sqrt{\frac{n}{L_{\min}(n)}}\log n\Big] = O(n^{-\gamma})$.*

i.e., the finish time on an n-ring, with random mobility, is $O(\sqrt{n}\log n)$ in expectation and with high probability.

We now prove that the finish time on a grid or line graph with *any* (possibly infection-state aware) external-infection spread strategy must be $\Omega(\sqrt{n})$, both in expectation and with high probability. This establishes that for ring graphs (or 1-dimensional grids), random external-infection is as good as any other form of controlled infection in an order-wise (up to a logarithm) sense. Furthermore, we use this theorem to introduce a general technique for obtaining lower bounds based on stochastic dominance via a parallel cluster-growing process. For ease of notation, we assume $||\bar{L}(t)||_1 \leq 1$ in this proof–in the next section, we obtain the more general bound (with dependence on $L_{\max}(n)$) for d-dimensional grids.

Theorem (Theorem 4: Lower bound for ring graphs)**.** *For the ring graph G_n with n nodes, and given that $||\bar{L}(t)||_1 \leq 1\,\forall t \geq 0$, there exists $c > 0$ independent of n such that for any spreading policy π,*

$$\mathbb{P}\left[T_\pi < c\sqrt{n}\right] = O\left(e^{-\Theta(1)\sqrt{n}}\right).$$

Moreover, we have

$$\inf_{\pi\in\Pi} \mathbb{E}[T_\pi] = \Omega\left(\sqrt{n}\right).$$

Proof: To keep the proof general, we use a parameter β for the intrinsic spreading rate over an edge (assumed to be 1 earlier). Along with the spreading process $(S^\pi(t))_{t\geq 0}$ induced by the policy π, consider a random process $(\tilde{S}(t))_{t\geq 0}$ described as follows:

(a) At all times t, $\tilde{S}(t)$ consists of an integer number $(\tilde{C}_t)$ of sets of points called *clusters*, where $(\tilde{C}_t)_{t\geq 0}$ is a Poisson process with intensity $L_{\max}(n) = 1$, and $\tilde{C}_0 = 1$ (the 1 denotes an 'initial' cluster in which static infection starts spreading).

(b) Once a new cluster is formed at some time s, it *grows*, *i.e.* adds points, following a Poisson process of intensity 2β (recall β is the intrinsic spreading rate for an edge in the graph).

Via a coupling argument, it can be shown that for *all* spreading strategies $\pi \in \Pi$, at all times $t \geq 0$, the total number of points in $\tilde{S}(t)$ (denoted by $\tilde{N}_t$) stochastically dominates that in $S^\pi(t)$ (informally, this is due to two reasons: first, that the rate of 'seeding' of new clusters by π is at most as fast as that in $\tilde{S}(\cdot)$; secondly, each cluster in $\tilde{S}(\cdot)$ grows independently and without interference from other existing clusters, as opposed to clusters that could 'merge' in the process $S^\pi(\cdot)$). Figure 1 graphically depicts the structure of the dominating process $\tilde{S}(\cdot)$.

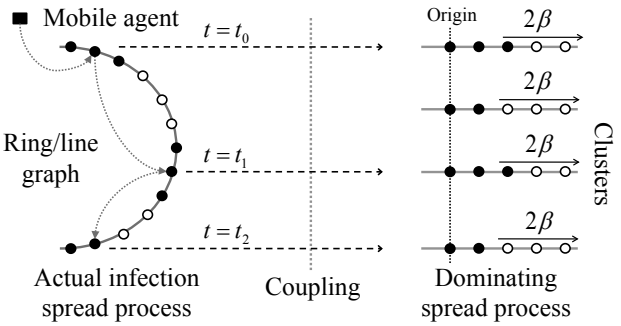

Fig. 1. Dominating the infection spread using independently growing clusters

Let $\tilde{T} \triangleq \inf\{t \geq 0 : \tilde{N}_t = n\}$ be the time when the number of points in $\tilde{S}(\cdot)$ first hits n. Owing to the stochastic dominance $\mathcal{N}(S^\pi(t)) \leq_{st} \tilde{N}_t$, we have that

$$\tilde{T} \leq_{st} T_\pi \;\forall \pi \in \Pi. \tag{10}$$

Knowing the way $\tilde{S}(\cdot)$ evolves, we can calculate $E[\tilde{N}_t]$:

$$\begin{aligned}\mathbb{E}[\tilde{N}_t] &= \mathbb{E}[\mathbb{E}[\tilde{N}_t|\tilde{C}_t]]\\ &= \sum_{k=0}^{\infty}\mathbb{P}(\tilde{C}_t = k)\mathbb{E}[\tilde{N}_t|\tilde{C}_t = k]\\ &= \sum_{k=0}^{\infty}\frac{e^{-t}t^k}{k!}\mathbb{E}[\tilde{N}_t|\tilde{C}_t = k].\end{aligned}$$

Since $\tilde{C}_t$ is a Poisson process, conditioned on $\{\tilde{C}_t = k\}$, the k cluster-creation instants are distributed uniformly on $[0, t]$. Let the times of these arrivals be $\tilde{T}_1, \ldots, \tilde{T}_k$; then $[\tilde{T}_i, t]$ is the time for which the ith cluster has been growing. Since every cluster grows at a rate of 2β, conditioned on $\{\tilde{C}_t = k\}$, the expected size of the ith cluster is $2\beta(t - \tilde{T}_i)$, $1 \leq i \leq k$. Also, the expected size of the '0-th' cluster at time t is $2\beta t$. Using $\mathbb{E}[\tilde{T}_i|\tilde{C}_t = k] = t/2$, we obtain

$$\begin{aligned}\mathbb{E}[\tilde{N}_t|\tilde{C}_t = k] &= 2\beta t + \sum_{i=1}^{k}\mathbb{E}[2\beta(t - \tilde{T}_i)|\tilde{C}_t = k]\\ &= \beta(k+2)t\\ \Rightarrow \mathbb{E}[\tilde{N}_t] &= \sum_{k=0}^{\infty}\frac{e^{-t}t^k}{k!}\mathbb{E}[\tilde{N}_t|\tilde{C}_t = k]\\ &= \sum_{k=0}^{\infty}\frac{e^{-t}t^k}{k!}\beta(k+2)t\\ &= \beta t^2 + 2\beta t.\end{aligned}$$

Hence, we have

$$P(\tilde{T} > t) = P(\tilde{N}_t < n) = 1 - P(\tilde{N}_t \geq n)$$
$$\geq 1 - \frac{\mathbb{E}[\tilde{N}_t]}{n} \geq 1 - \frac{\beta(t+1)^2}{n}.$$

by an application of Markov's inequality. This means that

$$\mathbb{E}[\tilde{T}] = \int_0^\infty \mathbb{P}(\tilde{T} > x)dx \geq \int_0^{\sqrt{\frac{n}{\beta}}-1} \mathbb{P}(\tilde{T} > x)dx$$
$$\geq \int_0^{\sqrt{\frac{n}{\beta}}-1} \left(1 - \frac{\beta(x+1)^2}{n}\right) dx$$
$$= \frac{2}{3}\sqrt{\frac{n}{\beta}} - 1 + \frac{\beta^2}{3n^2}$$
$$= \Theta(\sqrt{n}).$$

Together with (10), this forces $\inf_{\pi\in\Pi} \mathbb{E}[T_\pi] = \Omega(\sqrt{n})$, and the first part of the theorem is proved. For the second part, if we denote the size of the ith created cluster at time $s \geq T_i$ by $\tilde{X}_i(s)$, then for any time t we can write

$$\left(\bigcap_{i=0}^{2et} \{\tilde{X}_i(t+T_i) < 4e\beta t\}\right) \bigcap \{\tilde{C}_t < 2et\}$$
$$\subseteq \left(\bigcap_{i=0}^{\tilde{C}(t)} \{\tilde{X}_i(t+T_i) < 4e\beta t\}\right) \bigcap \{\tilde{C}_t < 2et\}$$
$$\subseteq \left(\bigcap_{i=0}^{\tilde{C}(t)} \{\tilde{X}_i(t) < 4e\beta t\}\right) \bigcap \{\tilde{C}_t < 2et\}$$
$$\subseteq \{\tilde{N}_t < 8\beta e^2 t^2\}.$$

Applying a standard Chernoff bound ($\mathbb{P}[Y \geq 2e\lambda] \leq (2e)^{-\lambda}$ for $Y \sim$ Poisson(λ)) to $\tilde{C}_t \sim$ Poisson(t) and $\tilde{X}_i(t+T_i) \sim$ Poisson($2\beta t$) above, we can write

$$\mathbb{P}[\tilde{N}_t \geq 8\beta e^2 t^2]$$
$$\leq \mathbb{P}[\tilde{C}_t \geq 2et] + \sum_{i=1}^{2et} \mathbb{P}[\tilde{X}_i(t+T_i) \geq 4e\beta t]$$
$$\leq (2e)^{-t} + 2et \cdot (2e)^{-2\beta t}$$
$$= O(e^{-t(1\wedge 2\beta)}).$$

In conclusion, using the stochastic dominance (10):

$$\mathbb{P}\left[T_\pi < \sqrt{\frac{n}{8\beta e^2}}\right] \leq \mathbb{P}\left[\tilde{T} < \sqrt{\frac{n}{8\beta e^2}}\right]$$
$$= \mathbb{P}\left[\tilde{N}_{\sqrt{\frac{n}{8\beta e^2}}} > n\right]$$
$$= O\left(e^{-\Theta(1)\sqrt{n}}\right).$$

■

B. d-Dimensional Grid Graphs

Extending the previous result, this section shows that the simple, state-oblivious random external-infection spreading strategy achieves the optimal order-wise finish time on d-dimensional grid networks for $d \geq 2$. For such a dimension d, the d-dimensional grid graph $G_n = (V_n, E_n)$ on n nodes is given by $V_n \triangleq \{1, 2, \ldots, n^{1/d}\}^d$, and $E_n \triangleq \{(x, y) \in V_n \times V_n : ||x - y||_1 = 1\}$.

Consider a partition of G_n into $(n/L_{\min})^{1/(d+1)}$ identical and contiguous 'sub-grids' $G_{n,i}$, $i = 1, \ldots, n^{1/(d+1)}$. By this, we mean that each $G_{n,i}$ is induced by a copy of $\{1, 2, \ldots, (n/L_{\min})^{1/(d+1)}\}^d$ (and thus has $(n/L_{\min})^{d/(d+1)}$ nodes). For instance, in the case of a planar $\sqrt{n} \times \sqrt{n}$ grid (with $L_{\min} = 1$), imagine tiling it horizontally and vertically with $\sqrt[3]{n}$ identical $\sqrt[3]{n} \times \sqrt[3]{n}$ sub-grids (Figure 2). With such a partition, an application of Theorem 1 shows that

Corollary (Corollary 2: Time for random spread on d-grids)**.** *For the random spread policy π_r on an n-node d-dimensional grid G_n, we have*

(a) $\mathbb{E}[T_{\pi_r}] = O\Big(\left(\frac{n}{L_{\min}(n)}\right)^{1/(d+1)} \log n\Big)$,

(b) *For any $\gamma > 0$ there exists $\alpha = \alpha(\gamma) > 0$ with* $\mathbb{P}\Big[T_{\pi_r} \geq \alpha\left(\frac{n}{L_{\min}(n)}\right)^{1/(d+1)} \log n\Big] = O(n^{-\gamma})$.

i.e., the finish time with random external-infection on a d-dimensional n-node grid is $O\left(\left(\frac{n}{L_{\min}(n)}\right)^{1/(d+1)} \log n\right)$ in expectation and with high probability.

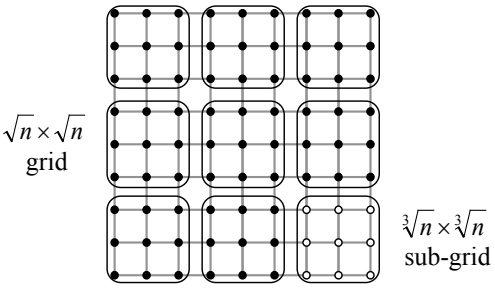

Fig. 2. Partitioning a planar grid into sub-grids

In what follows, we show that *any* external-infection spreading policy on a grid must take time $\Omega\left(\left(\frac{n}{L_{\max}(n)}\right)^{1/(d+1)}\right)$ to finish infecting all nodes with high probability, and consequently also in expectation. Barring a logarithmic factor, this shows that such a a random policy is as good as any other (possibly state-aware) policy for the class of grids. In order to derive this lower bound, we first need the following lemma from the theory of first-passage percolation [17], which essentially lets us control the extent to which infection on an infinite grid has spread at time t:

Lemma 2. *Let $(\tilde{Z}(t))_{t\geq 0} \in \{0,1\}^{\mathbb{Z}^d}$ represent a static/basic infection spread process on the infinite d-dimensional lattice $\mathbb{Z}^d$ starting at node $(0, 0, \ldots, 0)$ at time 0. Then, there exist positive constants l, c_3, c_4 such that for $t \geq 1$,*

$$\mathbb{P}[\mathcal{N}(\tilde{Z}(t)) > t^d l^d] \leq c_1 t^{2d} e^{-c_2\sqrt{t}}.$$

Proof of Lemma 2: Let

$$\tilde{B}(t) \triangleq \{v \in \mathbb{Z}^d : \tilde{Z}_v(t) = 1\} \subset \mathbb{Z}^d \ (\subset \mathbb{R}^d)$$

be the set of infected nodes at time t in $\tilde{Z}$. We use the following version of a result, from percolation on lattices

with exponentially distributed edge passage times, about the 'typical shape' of $\tilde{B}(t)$ [17]:

*(Theorem 2 in [17]) There exists a fixed (*i.e. *not depending on t) cube $B_0 = \left[-\frac{l}{2}, \frac{l}{2}\right]^d \subset \mathbb{R}^d$, and constants $c_1, c_2 > 0$, such that for $t \geq 1$,*

$$\mathbb{P}\left[\tilde{B}(t) \subset tB_0\right] \geq 1 - c_1 t^{2d} e^{-c_2\sqrt{t}}. \tag{11}$$

It follows from (11) that for $t \geq 1$,

$$\mathbb{P}[\mathcal{N}(\tilde{Z}(t)) > t^d l^d] = \mathbb{P}[|\tilde{B}(t)| > t^d l^d]$$
$$\leq \mathbb{P}[\tilde{B}(t) \nsubseteq tB_0] \leq c_1 t^{2d} e^{-c_2\sqrt{t}}.$$

■

Lemma 3 allows us to control the growth of individual seeded clusters of infected nodes; this is analogous to the use of the dominating spread process (growing at rate 2β) for line graphs. We thus obtain the following lower bound.

Theorem (Theorem 5: Lower bound: d-dimensional grids, bounded long-range virulence)**.** *Let G_n be a symmetric n-node d-dimensional grid graph. Suppose that $||\bar{L}(t)||_1 \leq L_{\max}(n) = \omega(n)$ for all $t \geq 0$. Then, there exist $c_1, c_2 > 0$, not depending on n, such that*

$$\mathbb{P}\left[T \leq c_1 \left(\frac{n}{L_{\max}(n)}\right)^{\frac{1}{d+1}}\right] = O\left(e^{-c_2\left(\frac{n}{L_{\max}(n)}\right)^{\frac{1}{2d+2}}}\right).$$

Furthermore, if $L_{\max}(n) = O(n^{1-\epsilon})$ for some $\epsilon \in (0, 1]$, then

$$\mathbb{E}[T] = \Omega\left(\left(\frac{n}{L_{\max}(n)}\right)^{\frac{1}{d+1}}\right).$$

Proof: Let us introduce a (dominating) counting process $(\tilde{S}(t))_{t\geq 0}$, described as follows:

- $\forall\, t \geq 0$, $\tilde{S}(t)$ consists of an integer number $(\tilde{C}_t)$ of sets of points called *clusters*, where $(\tilde{C}_t)_{t\geq 0}$ is a Poisson process with intensity $L_{\max}(n)$, and $\tilde{C}_0 = 1$ (denoting an 'initial' infected node).
- Each cluster grows as an independent copy of a static infection process on an exclusive infinite d-dimensional grid $\mathbb{Z}^d$ starting at $(0, 0, \ldots, 0)$.

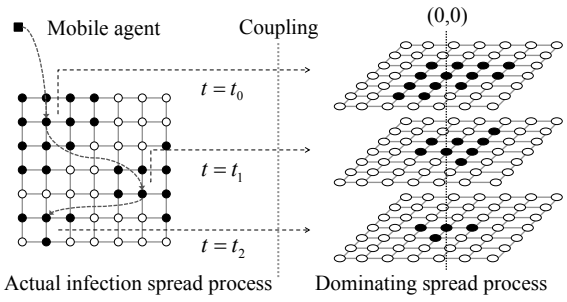

Fig. 3. The grid graph: Coupling infection spreading with mobility to a dominating 'cluster-growth' process

Note that in the process $\tilde{S}$, the growth of each cluster follows the natural static infection dynamics within a d-dimensional grid graph. Again, a standard coupling argument shows that at all times $t \geq 0$, the total number of points in $\tilde{S}(t)$ (denoted by $\tilde{N}_t$) stochastically dominates that in $S(t)$–this is essentially due to (a) cluster 'seeding' at the highest possible net exponential rate $L_{\max}(n)$, and (b) the absence of 'colliding' or multiple infections incident at any single node (Figure 3). Let $\tilde{T} \triangleq \inf\{t \geq 0 : \tilde{N}_t = n\}$ be the time when the number of points in $\tilde{S}(\cdot)$ first hits n. Then we have

$$\mathcal{N}(S(t)) \leq_{st} \tilde{N}_t \;\Rightarrow\; \tilde{T} \leq_{st} T. \tag{12}$$

Let us denote by $\tilde{X}_i(s)$ the size of the ith created cluster of $\tilde{S}(\cdot)$ at time $s \geq T_i$. Then, for $t \geq 0$, we have

$$\left(\bigcap_{i=0}^{2et}\{\tilde{X}_i(t + T_i) < t^d l^d\}\right) \bigcap \left(\{\tilde{C}_t < 2eL_{\max}(n)t\}\right)$$
$$\subseteq \{\tilde{N}_t < 2eL_{\max}(n) l^d t^{d+1}\},$$

Now each random variable $\tilde{X}_i(t + T_i)$ is distributed as the number of infected nodes in a static infection process on an infinite grid at time t. Thus, using Lemma 2 and a standard Chernoff bound for $\tilde{C}_t \sim \text{Poisson}(tL_{\max}(n))$, we can write

$$\mathbb{P}\left[\tilde{N}_t \geq (2eL_{\max}(n) l^d) t^{d+1}\right]$$
$$\leq \mathbb{P}\left[\tilde{C}_t \geq 2eL_{\max}(n)t\right] + \sum_{i=1}^{2eL_{\max}(n)t} \mathbb{P}\left[\tilde{X}_i(t + T_i) \geq t^d l^d\right]$$
$$\leq (2e)^{-L_{\max}(n)t} + 2eL_{\max}(n)t \cdot c_3 t^{2d} e^{-c_4\sqrt{t}}$$
$$= O(L_{\max}(n) e^{-c_4\sqrt{t}}).$$

With the stochastic dominance (12), this forces

$$\mathbb{P}\left[T \leq \left(\frac{n}{2eL_{\max}(n)l^d}\right)^{1/(d+1)}\right]$$
$$\leq \mathbb{P}\left[\tilde{T} \leq \left(\frac{n}{2eL_{\max}(n)l^d}\right)^{1/(d+1)}\right]$$
$$= \mathbb{P}\left[\tilde{N}_{\left(\frac{n}{2eL_{\max}(n)l^d}\right)^{1/(d+1)}} \geq n\right]$$
$$= O\left(e^{-c_2\left(\frac{n}{L_{\max}(n)}\right)^{1/(2d+2)}}\right), \tag{13}$$

for the appropriate c_2, establishing the first part of the theorem. To see how this implies the second part, note that the estimate (13), together with the fact that $L_{\max}(n) = O(n^{1-\epsilon})$ and the Borel-Cantelli lemma, gives us

$$\mathbb{P}\left[\tilde{T} \leq \left(\frac{n}{2eL_{\max}(n)l^d}\right)^{1/(d+1)} \text{ for finitely many } n\right] = 1,$$
$$\Rightarrow \liminf_{n\to\infty} \frac{\tilde{T}}{(n/L_{\max}(n))^{1/(d+1)}} \overset{a.s.}{\geq} c_4 \triangleq \frac{1}{(2el^d)^{1/(d+1)}} > 0$$

By Fatou's lemma,

$$\liminf_{n\to\infty} \mathbb{E}\left[\frac{\tilde{T}}{(n/L_{\max}(n))^{1/(d+1)}}\right]$$
$$\geq \mathbb{E}\left[\liminf_{n\to\infty} \frac{\tilde{T}}{(n/L_{\max}(n))^{1/(d+1)}}\right] \geq c_4 > 0.$$

This shows that $\mathbb{E}[T] \geq \mathbb{E}[\tilde{T}] = \Omega\left((n/L_{\max}(n))^{1/(d+1)}\right)$, and concludes the proof of the theorem. ∎

C. Geometric Random Graphs

We finally prove the upper and lower bounds for the *Geometric Random Graph (RGG)*. Recall that an RGG is a family of random graphs wherein n points (*i.e.* nodes) are picked i.i.d. uniformly in $[0,1] \times [0,1]$. Two nodes x, y are connected by an edge iff $||x - y|| \leq r_n$, where r_n is often called the *coverage radius*. The RGG $G_n = G_n(r_n)$ consists of the n nodes and edges as above.

It is known that when the coverage radius r_n is above a critical threshold of $\sqrt{\log n/\pi}$, the RGG is connected with high probability [35]. In this section, we state and prove two results that show that random spreading on RGGs in this critical connectivity regime is optimal upto logarithmic factors. First, we show with high probability that random spreading finishes in time $O(\sqrt[3]{n}\log n)$, and follow it up with a converse result that says that no other policy can better this order (up to the logarithmic factor) with significant probability. This directly parallels the earlier results about finish times on 2-dimensional grids, where random mobile spread exhibits the same optimal order of growth.

Theorem (Theorem 6: Time for random spread on RGGs)**.** *For the planar random geometric graph $G_n(r_n)$, if $r_n \geq \sqrt{\frac{5\log n}{n}}$, then there exists $\alpha > 0$ such that*

$$\lim_{n\to\infty} \mathbb{P}[T_{\pi_r} \geq \alpha \sqrt[3]{n/L_{\min}(n)}\log n] = 0.$$

Proof: Divide the unit square $[0,1]\times[0,1]$ into square *tiles* of side length $r_n/\sqrt{5}$ each; there are thus $5/r_n^2$ such tiles, say $k_1, \ldots, k_{5/r_n^2}$. If n points are thrown uniformly randomly into $[0,1] \times [0,1]$, then, with $\mathcal{E}$ denoting the event that some tile is empty,

$$\begin{aligned}
\mathbb{P}[\mathcal{E}] &\leq \frac{5}{r_n^2}\mathbb{P}[\text{tile 1 empty}] \\
&= \frac{5}{r_n^2}\left(1 - \frac{r_n^2}{5}\right)^n \leq \frac{5}{r_n^2}\exp\left(-\frac{nr_n^2}{5}\right) \\
&\leq \frac{n}{\log n}\exp(-\log n) \\
&= \frac{1}{\log n} \overset{n\to\infty}{\longrightarrow} 0. \qquad (14)
\end{aligned}$$

By construction, note that the maximum distance between points in two horizontally or vertically adjacent tiles is exactly r_n. Hence, two nodes in horizontally or vertically adjacent tiles are always connected by an edge. Also, a node in a tile is not connected to any node in a tile at least three tiles away in either dimension.

If we now divide $[0,1] \times [0,1]$ into (bigger) square *chunks* of side length $1/\sqrt[6]{nL_{\min}(n)^2}$ each, there are $\sqrt[3]{\frac{n}{L_{\min}(n)}}$ such square chunks, each containing a $\frac{\sqrt{5}}{r_n\sqrt[6]{n}} \times \frac{\sqrt{5}}{r_n\sqrt[6]{n}}$ grid of square tiles. In the case where no tile is empty, it follows from the arguments in the preceding paragraph that the diameter of the subgraph induced within each chunk is

$$\begin{aligned}
O\left(\frac{1/\sqrt[6]{nL_{\min}(n)^2}}{r_n}\right) &= O\left(\frac{1}{r_n\sqrt[6]{nL_{\min}(n)^2}}\right) \\
&= O\left(\frac{\sqrt[3]{\frac{n}{L_{\min}(n)}}}{\sqrt{\log n}}\right)
\end{aligned}$$

since $r_n \geq \sqrt{5\log n}$. An application of Theorem 1 in this case shows that $\mathbb{E}[T_{\pi_r}|\mathcal{E}] = O(\sqrt[3]{n/L_{\min}(n)}\log n)$, and for some $\alpha, \gamma > 0$, $\mathbb{P}\left[T_{\pi_r} \geq \alpha\sqrt[3]{n/L_{\min}(n)}\log n \mid \mathcal{E}\right] = O(n^{-\gamma})$. Using (14), we conclude that

$$\mathbb{P}[T_{\pi_r} \geq \alpha\sqrt[3]{n/L_{\min}(n)}\log n] = O\left(\frac{1}{\log n}\right) \overset{n\to\infty}{\longrightarrow} 0.$$

∎

Consider an infinite planar grid with additional one-hop diagonal edges, i.e. $G = (V, E)$ where $V = \mathbb{Z}^2$, $E = \{(x, y) \in \mathbb{Z}^2 : ||x - y||_\infty \leq 1\}$. Let an infection process $(S(t))_{t\geq 0}$ start from $0 \in \mathbb{Z}^2$ at time 0 according to the standard static spread dynamics, i.e. with each edge propagating infection at an exponential rate μ, and let $I(t)$ denote the set of infected nodes at time t. The following key lemma helps control the size of $I(t)$, *i.e.* the extent of infection at time t:

Lemma 3. *There exists $c_1 > 0$ such that for any $c_2 > 0$ and t large enough,*

$$\begin{aligned}
&\mathbb{P}\left[\exists x \in I(t) : ||x||_\infty \geq (c_1\mu + c_2)t\right] = \\
&\quad O\left((c_1\mu + c_2)t \cdot e^{-c_2 t}\right).
\end{aligned}$$

Proof:

$$\begin{aligned}
&\mathbb{P}[\exists x \in I(t) : ||x||_\infty \geq ct] \\
&\quad \leq \mathbb{P}[\exists v \in \mathbb{Z}^2 : ||v||_\infty = \lfloor ct\rfloor, T(v) \leq t] \\
&\quad \leq \sum_{v\in\mathbb{Z}^2 : ||v||_\infty = \lfloor ct\rfloor} \mathbb{P}[\exists \text{ a path } r : 0 \to v, T(r) \leq t].
\end{aligned}$$

Observe that for any v with $||v||_\infty = \lfloor ct\rfloor$ and any path of edges r from 0 to v, there must exist $\lfloor ct\rfloor + 1$ nodes $v_0 = 0, v_1, \ldots, v_{\lfloor ct\rfloor}$ on the path r such that $||v_i||_\infty \leq \lfloor ct\rfloor$ and $||v_{i+1} - v_i||_\infty = 1$. Indeed, each edge on a path can increase the $||\cdot||_\infty$ distance from 0 by at most 1. Therefore, continuing the above chain of inequalities, the sought probability is bounded above by

$$\begin{aligned}
&\sum_{\{v:||v||_\infty = \lfloor ct\rfloor\}} \sum_{v_0,\ldots,v_{\lfloor ct\rfloor}} \mathbb{P}\left[\exists \text{ a path } r : 0 \to v \text{ passing}\right. \\
&\left.\text{successively through the } v_i, T(r) \leq t\right] \\
&\leq \sum_{\{v:||v||_\infty = \lfloor ct\rfloor\}} \sum_{v_0,\ldots,v_{\lfloor ct\rfloor}} \mathbb{P}\left[\exists \text{ a path } r \text{ passing}\right. \\
&\left.\text{successively through the } v_i, \sum_{i=0}^{\lfloor ct\rfloor - 1} T(v_i, v_{i+1}) \leq t\right] \\
&\leq \sum_{\{v:||v||_\infty = \lfloor ct\rfloor\}} \sum_{v_0,\ldots,v_{\lfloor ct\rfloor}} \mathbb{P}\left[\sum_{i=0}^{\lfloor ct\rfloor - 1} T(v_i, v_{i+1}) \leq t\right], \qquad (15)
\end{aligned}$$

where the second sum runs throughout over all v_i with $v_0 = 0$,

$||v_i||_\infty \le \lfloor ct \rfloor$ and $||v_{i+1} - v_i||_\infty = 1$, and $T(x,y)$ represents the infection passage time from node x to node y. Letting $T'(v_i, v_{i+1})$ be random variables identically distributed as $T(v_i, v_{i+1})$ but *independent* for $i = 1, \ldots, \lfloor ct \rfloor - 1$, we can write, for $\psi > 0$,

$$\begin{aligned}
&\sum_{v_0,\ldots,v_{\lfloor ct \rfloor}} \mathbb{P}\left[\sum_{i=0}^{\lfloor ct \rfloor - 1} T(v_i, v_{i+1}) \le t\right] \\
&\quad = \sum_{v_0,\ldots,v_{\lfloor ct \rfloor}} \mathbb{P}\left[\sum_{i=0}^{\lfloor ct \rfloor - 1} T'(v_i, v_{i+1}) \le t\right] \\
&\quad \le \sum_{v_0,\ldots,v_{\lfloor ct \rfloor}} e^{\psi t} \prod_{i=0}^{\lfloor ct \rfloor - 1} \mathbb{E}\left[e^{-\psi T'(v_i, v_{i+1})}\right] \\
&\quad = e^{\psi t} \sum_{v_0,\ldots,v_{\lfloor ct \rfloor}} \prod_{i=0}^{\lfloor ct \rfloor - 1} \mathbb{E}\left[e^{-\psi T'(v_i, v_{i+1})}\right] \\
&\quad = e^{\psi t} \left(\sum_{\{u : ||u||_\infty = 1\}} \mathbb{E}\left[e^{-\psi T'(0,u)}\right]\right)^{\lfloor ct \rfloor}.
\end{aligned}$$

In the last step of the above display, we have successively summed over $v_{\lfloor ct \rfloor}, v_{\lfloor ct \rfloor - 1}, \ldots, v_0$, and have used the fact that infection spread times are translation-invariant, *i.e.* for any $x, y, a \in \mathbb{Z}^2$,

$$\begin{aligned}
T'(x,y) &\stackrel{d}{=} T(x,y) \stackrel{d}{=} T(x+a, y+a) \\
&\stackrel{d}{=} T'(x+a, y+a).
\end{aligned}$$

For any $u \in \mathbb{Z}^2$ such that u is a neighbour of 0 (*i.e.* $||u||_\infty = 1$), we must have $T(0,u) \ge \min_{w : ||w||_\infty = 1} t((0,w))$, where $t(e) \sim \text{Exp}(\mu)$ is the travel time of the infection across edge $e \in E$. Since the number of neighbours of 0 in G is exactly 8 (4 up-down/left-right and 4 diagonal), $T(0,u)$ stochastically dominates an exponential random variable with parameter 8μ. Thus,

$$\mathbb{E}\left[e^{-\psi T'(u,v)}\right] \le \mathbb{E}\left[e^{-\psi \hat{T}}\right] = \left(1 + \frac{\psi}{8\mu}\right)^{-1},$$

(where $\hat{T} \sim \text{Exp}(8\mu)$)

$$\begin{aligned}
&\Rightarrow e^{\psi t} \left(\sum_{\{u : ||u||_\infty = 1\}} \mathbb{E}\left[e^{-\psi T'(0,u)}\right]\right)^{\lfloor ct \rfloor} \\
&\quad \le e^{\psi t} \left(8\left(1 + \frac{\psi}{8\mu}\right)^{-1}\right)^{\lfloor ct \rfloor}. \qquad (16)
\end{aligned}$$

Setting $\psi = 8\mu(8e-1)$ so that $8(1+\psi/\mu)^{-1} = e^{-1}$, (16) becomes

$$\begin{aligned}
&e^{\psi t}\left(\sum_{\{u : ||u||_\infty = 1\}} \mathbb{E}\left[e^{-\psi T'(0,u)}\right]\right)^{\lfloor ct \rfloor} \\
&\qquad \le e^{8\mu(8e-1)t} \cdot e^{-ct+1}.
\end{aligned}$$

Finally, letting $c_1 = 8(8e-1)$ and $c = c_1\mu + c_2$, we obtain the desired result from (15) and the above:

$$\begin{aligned}
&\sum_{\{v : ||v||_\infty = \lfloor ct \rfloor\}} \sum_{v_0,\ldots,v_{\lfloor ct \rfloor}} \mathbb{P}\left[\sum_{i=0}^{\lfloor ct \rfloor - 1} T(v_i, v_{i+1}) \le t\right] \\
&\qquad \le |\{v : ||v||_\infty = \lfloor ct \rfloor\}| \cdot e^{-c_2 t + 1} \\
&\qquad \le (4ct) \cdot e^{-c_2 t + 1} \\
&\qquad = O\left((c_1\mu + c_2)t \cdot e^{-c_2 t}\right).
\end{aligned}$$

■

Using Lemma 3, we can finally state our converse result for the geometric random graph:

Theorem (Theorem 7: Lower bound for RGGs)**.** *For the planar geometric random graph G_n with $r_n = O(\sqrt{\log n / n})$ and any spreading policy π with $L_{\max}(n) = O(n^{1-\epsilon})$ for some $\epsilon \in (0,1]$, $\exists\ \beta > 0$ s.t. $\lim_{n\to\infty} \mathbb{P}\left[T_\pi \ge \beta \frac{\sqrt[3]{n / L_{\max}(n)}}{\log^{4/3} n}\right] = 1$.*

Proof sketch: The method of approach is along the lines of that used to prove Theorem 5 along with certain geometric considerations for the case of the random geometric graph. We introduce a spreading process that spreads 'faster' than π, and show using Lemma 3 that even this process must take at least the claimed amount of time to spread. For ease of exposition, we break the proof into two steps:

Step 1: Divide the unit square $[0,1] \times [0,1]$ row and column-wise into $r_n \times r_n$ *tiles*; there are thus $1/r_n^2$ tiles, say $k_1, \ldots, k_{1/r_n^2}$. By standard balls-and-bins arguments, with the n nodes thrown randomly into $n/\log n$ tiles, each tile receives a maximum of $O(\log n)$ nodes with probability $1 - o(1)$.

Step 2: Within the event in step 1, we introduce the following associated spreading process which, via coupling arguments, can be shown to dominate the spread due to π at each time t: first, take each tile to be the vertex of a square grid where adjacent diagonals are connected. Also, set the rate of infection spread on every edge to be $\text{Exp}(\mu \log^2 n)$. This effectively upper-bounds the best rate of spread among neighbouring tiles. Create a dominating process by creating non-interfering clusters at a Poisson rate 1, with each cluster growing independently on an infinite square grid with diagonal edges and the above spread rate. Lemma 3 shows that w.h.p., by time t, $O(t)$ clusters are formed, and each cluster has at most $O(t^2 \log^4 n)$ nodes. Thus it takes at least $O\left(\frac{\sqrt[3]{n}}{\log^{4/3} n}\right)$ time for spreading to finish w.h.p. ■

VI. Conclusion

We have modelled and analyzed the spread of epidemic processes on graphs when assisted by external agents. For general graphs, we have provided upper bounds on the spreading time due to external-infection with bounded virulence for random and greedy infection policies; these bounds are in terms of the diameter and the conductance of the graph. On the other hand, for certain spatially-constrained graphs such as grids and the geometric random graph, we have derived corresponding lower bounds: these indicate that random external-infection spreading is order-optimal up to logarithmic factors (and greedy is order-optimal) in such scenarios. Finally, we have

discussed applications of our result to graphs with long-range edges and/or mobile agents.

REFERENCES

[1] A. Gopalan, S. Banerjee, A. K. Das, and S. Shakkottai, "Random mobility and the spread of infection," in *IEEE INFOCOM*, pp. 999–1007, 2011.

[2] F. G. Ball, "Stochastic multitype epidemics in a community of households: Estimation of threshold parameter R* and secure vaccination coverage," *Biometrika*, vol. 91, no. 2, pp. 345–362, 2004.

[3] B. Wang, K. Aihara, and B. J. Kim, "Immunization of geographical networks," in *Complex (2)*, pp. 2388–2395, 2009.

[4] R. M. Anderson and R. M. May, *Infectious Diseases of Humans Dynamics and Control*. Oxford University Press, 1992.

[5] E. M. Rogers, *Diffusion of Innovations, 5th Edition*. Free Press, original ed., August 2003.

[6] M. Granovetter, "The Strength of Weak Ties," *The American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.

[7] J. O. Kephart and S. R. White, "Directed-graph epidemiological models of computer viruses," in *IEEE Symposium on Security and Privacy*, pp. 343–361, 1991.

[8] G. Kossinets, J. Kleinberg, and D. Watts, "The structure of information pathways in a social communication network," in *KDD '08: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, (New York, NY, USA), pp. 435–443, ACM, 2008.

[9] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *KDD '03: Proc. 9th ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, pp. 137–146, ACM, 2003.

[10] R. Pastor-Satorras and A. Vespignani, "Epidemic dynamics in finite size scale-free networks," *Phys. Rev. E*, vol. 65, p. 035108, Mar 2002.

[11] A. J. Ganesh, L. Massoulié, and D. F. Towsley, "The effect of network topology on the spread of epidemics," in *IEEE INFOCOM*, pp. 1455–1466, 2005.

[12] B. Pittel, "On spreading a rumor," *SIAM J. Appl. Math.*, vol. 47, no. 1, pp. 213–223, 1987.

[13] S. Sanghavi, B. Hajek, and L. Massoulié, "Gossiping with multiple messages," in *IEEE INFOCOM*, pp. 2135–2143, 2007.

[14] M. Grossglauser and D. Tse, "Mobility Increases the Capacity of Ad Hoc Wireless Networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, p. 477, 2002.

[15] A. D. Sarwate and A. G. Dimakis, "The impact of mobility on gossip algorithms," in *IEEE INFOCOM*, pp. 2088–2096, 2009.

[16] R. Durrett and X.-F. Liu, "The contact process on a finite set," *Annals of Probability*, vol. 16, pp. 1158–1173, 1988.

[17] H. Kesten, "On the speed of convergence in first-passage percolation," *The Annals of Applied Probability*, vol. 3, no. 2, pp. 296–338, 1993.

[18] F. Ball, D. Mollison, and G. Scalia-Tomba, "Epidemics with two levels of mixing.," *Ann. Appl. Probab.*, vol. 7, no. 1, pp. 46–89, 1997.

[19] P. Wang, M. C. González, C. A. Hidalgo, and A.-L. Barabasi, "Understanding the spreading patterns of mobile phone viruses," *Science*, vol. 324, pp. 1071–1076, May 2009.

[20] J. Cheng, S. H. Wong, H. Yang, and S. Lu, "Smartsiren: virus detection and alert for smartphones," in *MobiSys '07: Proceedings of the 5th international conference on Mobile systems, applications and services*, (New York, NY, USA), pp. 258–271, ACM, 2007.

[21] J. W. Mickens and B. D. Noble, "Modeling epidemic spreading in mobile environments," in *WiSe '05: Proceedings of the 4th ACM Workshop on Wireless Security*, (New York, NY, USA), pp. 77–86, ACM, 2005.

[22] J. Su, K. K. W. Chan, A. G. Miklas, K. Po, A. Akhavan, S. Saroiu, E. de Lara, and A. Goel, "A preliminary investigation of worm infections in a Bluetooth environment," in *WORM '06: Proceedings of the 4th ACM workshop on Recurring malcode*, (New York, NY, USA), pp. 9–16, ACM, 2006.

[23] J. Kleinberg, "The Wireless Epidemic," *Nature*, vol. 449, pp. 287–288, September 2007.

[24] S. Riley, "Large-scale spatial-transmission models of infectious disease," *Science*, vol. 318, no. 5847, 2007.

[25] E. Kaplan, D. L. Craft, and L. M. Wein, "Analyzing bioterror response logistics: the case of smallpox," *Math Biosci*, vol. 185, p. 33072, September 2003.

[26] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, "The role of the airline transportation network in the prediction and predictability of global epidemics," *Proceedings of the National Academy of Sciences*, vol. 103, no. 7, pp. 2015–2020, 2006.

[27] D. Kempe, J. Kleinberg, and A. Demers, "Spatial gossip and resource location protocols," *J. ACM*, vol. 51, no. 6, pp. 943–967, 2004.

[28] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," *arXiv*, Jun 2010.

[29] D. Shah, "Gossip algorithms," *Found. Trends Netw.*, vol. 3, no. 1, pp. 1–125, 2009.

[30] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 21484–21489, December 2009.

[31] N. Alon, "Transmitting in the n-dimensional cube," *Discrete Appl. Math.*, vol. 37-38, pp. 9–11, Jul 1992.

[32] C. Martel and V. Nguyen, "Analyzing Kleinberg's (and other) small-world models," in *PODC '04: Proc. 23rd annual ACM symposium on Principles of Distributed Computing*, pp. 179–188, ACM, 2004.

[33] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, June 1998.

[34] P. Bremaud, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer-Verlag New York Inc., corrected ed., February 2001.

[35] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," in *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming* (W. M. McEneany, G. Yin, and Q. Zhang, eds.), pp. 547–566, Boston: Birkhauser, 1998.