Epidemics with SI Dynamics and External Agents (Original Paper - Bounds and Uses)

Joel Anto Paul, Anway Deshpande, Tamojeet Roychowdhury

***Abstract*—Graphs with nodes representing objects and edges representing their interaction are frequently used to model the spread of several phenomena, such as human/animal/plant diseases, computer viruses across networks, information through media and advertisements, all collectively termed as epidemics. The SI (susceptible-infected) dynamic is often a good model for many of these spreading processes, specially in cases where recovery is not possible. The metric of interest here is how much time it takes for the entire graph (community) to get infected, known as the spreading time. This paper takes the SI model and adds external infecting agents to it, that can infect nodes that have no neighbours. They derive upper bounds on the spreading time in both concentration and expectation, for general graphs with specific external infection policies. These depend on the graph properties like the diameter, subgraph-partitioning and conductance. Further, they derive lower bounds on the spreading time for specific graph topologies and general (possibly greedy) infecting policies, again utilizing the graph structures and stochastic dominance. These theorems result in some surprising consequences on the efficacy of mobile external agents randomly vs greedily infecting nodes through the graph and the resultant spreading times.**

I. Introduction

A set of several interacting objects, alternatively called a community, can be modelled as a graph $G(V, E), |V| = n$, where each vertex represents an object in the community and there exists an edge between two nodes if those two interact with each other. In the simplest scenario, these edges are undirected and unweighted and are generally arbitrary in number.

In this community represented by a graph, any *spreading phenomena*, collectively called an epidemic, can be modelled as originating at a particular node, and then spreading through via node-node interaction, that is, via edges. Since every interaction resulting in the spread of the phenomena is unlikely, this spread is modelled as a stochastic process where the time taken for one infected node to infect the next is often exponentially distributed. For processes where getting infected once leads to permanent contraction of the disease, the simplest model consists of all nodes going from a susceptible state (S) to an infected state (I) and never recovering after that, henceforth called the *SI model*. There are other more complicated models where recovery is possible, resulting in the node becoming resistant to future infections (SIR) or susceptible again (SIS). The SI model is obviously not good for modelling a lot of human diseases where recovery is possible, but it can model well several other systems such as

- *Plant diseases* - Since plants lack an immune system, any epidemic that spreads through a crop field or a forest (a crowded ecosystem) is likely to permanently infect the plants and is typically well modelled by SI dynamics before a pesticide is designed to combat it.
- *Computer viruses* - Modern antivirus has evolved today to thwart many malware attacks on the host computer itself before spreading starts. However, a particularly resistant computer virus that can not easily be cleaned by existing antivirus software, is likely to rapidly multiply and spread across systems before a remedy is developed and tested to work successfully. Since recovery is not immediately available, the SI model works here.
- *Invasive species and gene transfer* - If an invasive species (birds/weeds/insects) spreads and begins reproducing in a given geographical area or locality, it is usually very difficult to curb or eradicate their presence, multiplication and further spread through air or other vector channels. Similarly, an important component of evolution is gene transfer between different locations. Here again once a gene is introduced into a population, it gradually grows and is never removed again. It can subsequently spread to other nearby areas,
- *Social trends/influence* - Various models may work nearly equally well for this scenario; however once a social trend has set in, more and more people are likely to copy the style and result in it spreading to more people, and the persistence of that trend is often slow enough in wearing out to make SI model a good fit for the situation.

The paper looks at infections spreading via SI dynamics across connected nodes (henceforth called *intrinsic infection*), as well as induced externally at random healthy nodes via infectious agents, unconstrained by graph parameters (henceforth called *external infection*). The original paper uses these two important distributions:

(a) The time between two consecutive nodes getting infected via external infection is exponentially distributed with node-dependent rates $L_i, i \in \{1, 2 \dots n\}$

(b) The time between the infection of one node to the infection of next node on a connected path via intrinsic infection is exponentially distributed with rate 1.

A few physical scenarios that can be modelled well by external agents are

- *Long-distance travellers* - In the spread of invasive species, gene transfer or human diseases without recovery, spread is often limited by geography and tightly connected subgraphs. There are occasionally however

tamojeet@iitb.ac.in

some long-distance travellers that can spread the epidemic if they are carrying the infection. Such a scenario does not need a permanent edge as these travels are typically limited in number, but are captured well as external agents
- *Targeted advertising* - Fast spreading times are desirable for product advertising. Word-of-mouth or other means often serve as the intrinsic spreading process here, whereas the external infection may correspond to targeting particular shopping outlets, selective discounts and so on, the efficacies and consequently policies of which can be designed and informed from the results derived in the paper
- *Short and long-range worms* - Computer/mobile worms often spread across short networks such as Bluetooth or a shared local network which corresponds to the intrinsic spread, and sometimes across different local networks via long-distance sharing such as mails or SMS.

Recall that the external virulence/infectivity for i^{th} node is taken to be L_i. The paper imposes a constraint on the total virulence $\sum_{i=1}^{n} L_i(t) = L(t)$. This is taken to have maximum and minimum limits L_{min} and L_{max}.

The paper derives two kinds of bounds on the spreading time for two classes of the problem:

1) *Upper bounds* - These are given for general graphs and specific policies. The graph is partitioned into subgraphs, and the properties of the subgraphs (number, maximum/minimum of sizes, diameters and conductance) are used to derive the upper bounds. The two specific policies analyzed are the *random spreading*, where L_i is same for all nodes, and *greedy spreading*, where the external agent is omniscient and selectively infects one node in each subgraph in sequence.
2) *Lower bounds* - These are given for general external spreading policies, which could possibly be adversarial, and specific graphs which includes rings/linear graphs, d-dimensional grids and random geometric graphs. Lower bounds on the spreading time for a fixed n number of nodes is derived by upper bounding the maximum number of nodes that can be infected in a fixed time t. This is primarily achieved by creating independent clusters for each externally infected node.

The last section discusses some specific manifestations of these external agents and results pertaining to their role in decreasing the spreading time using mainly order-wise bounds.
Defnition of stochastic dominance: A random variable A is said to stochastically dominate another random variable B if for all real x, $\mathbb{P}(A > x) > \mathbb{P}(B > x)$ and denoted as $A \geq_{st} B$.

II. Graph Properties and Notation

The following graph properties play an important role in the bounds. Their relevant notation and qualitative physical properties are outlined below:

- Number of subgraphs $g(\Pi)$: If we want to keep the cluster sizes roughly same, this directly scales with the graph size n. Owing to the way we divide the spreading time into two phases for the proofs, this results in the external spreading taking a longer time if $g(\Pi)$ is larger.
- Maximum and minimum size of subgraphs $s(\Pi)$: If the number of subgraphs remains roughly constant, then the size scales with the size of he graph, and the time taken for intrinsic infection increases. Combined with the above quantity, this gives us the obvious fact that as size of graph increases, the spreading time must also increase - the order of that increase is of interest to us and derived in the following theorems.
- Maximum diameter of any subgraph $d(\Pi)$: The *distance* between two nodes is given by the path with least number of edges to be crossed. The longest of these shortest paths in a subgraph is called the diameter of that subgraph. Roughly, as the diameter increases, in the worst-case scenario, the infection has to travel more path to reach the end nodes, and the intrinsic infection time increases.
- Conductance of a graph Ψ: If S is a subset of the set of all vertices V in a graph, containing less than half the vertices, then the conductance is given by
$$\Psi = \inf_S \frac{E(S, V \backslash S)}{|S|}$$
which physically translates to the minimum number of edges that connect one subset of the graph to the remaining graph, scaled down by the number of vertices in the smaller subgraph. Higher conductance means a well-connected graph, leading to a faster spread of the infection in the subgraph.

We now look at the specific theorems on graphs and policies derived using constraints on the above properties.

III. Upper Bounds for General Graphs under Specific Policies

Theorem 1: (Upper bound, Diameter version). For a graph G and any partition Π(G) connected by g(Π) subgraphs, each with size at least s(Π) and diameter at least d(Π), the total spreading time satisfies the following bounds:

(a)
$$\mathbb{E}[T] \leq h(\Pi).(log(n) + 1)$$
where h(Π) = max($\frac{n}{s(\Pi)L_{min}}$,d(Π))

(b) If g(Π) $\leq cn^{\delta}$ for some constants c, $\delta > 0$, then for any $\gamma > 0$,
$$\mathbb{P}[T \geq \kappa h(\Pi) logn] \leq c' n^{-\gamma}$$
where $\kappa > 1 + \frac{\gamma}{\delta}$ and c' = $2c^{-\kappa+1}$

Proof of Theorem 1: The authors create a new slower spreading process that involves two spreading times T_1 and T_2 which are defined as:

- T_1 is the time by which at least one node in each subgraph of the partition Π is infected. We assume infection in this stage to spread via external agents alone, and not via intrinsic spread.
- Once infection has started in each subgraph, T_2 is the additional time by which all nodes in each subgraph has been infected. No external infection is assumed to operate in this phase of spread.

Clearly, T_1 + T_2 stochastically dominates the original spreading time T. Now to bound T_1, we can use the memorylessness

property of exponential random variable so T_1 is stochastically bounded by maximum of these i.i.d random variables, each with rate atleast $\frac{L_{min}s}{n}$. Let $T_{11}, T_{12}, ..., T_{1g}$ be spreading times for each subgraphs. Now all these are exponential random variables with rate atleast $\frac{L_{min}s}{n}$, So,

$$\mathbb{E}[T_1 i] \leq \frac{1}{\frac{L_{min}s}{n}} = \frac{n}{sL_{min}} \text{ for all i = 1 to g}$$

So, now define a new random variable T_{1max} = max$\{T_{11}, T_{12}, ..., T_{1g}\}$. So, $\mathbb{P}[T_{1max} \leq x] = \mathbb{P}[T_{11} \leq x$ and $T_{12} \leq x$ and... $T_{1n} \leq x] = (\mathbb{P}[T_{11} \leq x])^n = (1 - e^{-\frac{n}{L_{min}s_{min}}})^n$. So,

$$\begin{aligned}\mathbb{E}[T_1] &\leq \int_{x=0}^{\infty} \mathbb{P}[T_{1max} > x]dx \\ &= \int_{x=0}^{\infty} (1 - (1 - e^{-\frac{n}{L_{min}s_{min}}})^n)dx \\ &= \sum_{i=0}^{g} \frac{n}{iL_{min}s_{min}} \\ &\leq \frac{n(\log g + 1)}{sL_{min}}\end{aligned}$$

Using union bounds on tails of g i.i.d exponential random variables, for any $\kappa > 0$, we get

$$\mathbb{P}[T_1 \geq \kappa \frac{n \log g}{sL_{min}}] \leq g e^{-(\frac{sL_{min}}{n} \frac{\kappa n \log g}{sL_{min}})}$$

Now, to bound T_2, we can upper bound the expectation by the term : length of the longest path $\times$ maximum time to spread across an edge. So, we get

$$\mathbb{E}[T_2] \leq d.(\log n + 1)$$

Again, using union bounds on tail probability of T_2, for any $\kappa > 0$

$$\mathbb{P}[T_2 \geq \kappa d. \log n] \leq n^{-\kappa+1}$$

Combining the results, we get the required bounds.

Theorem 2: (Upper bound for Greedy Subgraph Infection(GSI) Policy). For a graph G and any partition Π(G) connected by g(Π) subgraphs, each with size at least s(Π) and diameter at max d(Π). Also, $d(\Pi) \geq \log n$. Then for the Greedy subgraph infection policy:

$$\mathbb{E}[T] \leq max(\frac{g(\Pi}{L_{min}}, 4d(\Pi))$$

Proof of Theorem 2: The proof follows the same line, but now phase 1 consists of sequential injecting of infection in each subgraph, so T_1 now corresponding to sum of i.i.d exponential random variables rather than maximum of them, so we get the standard concentration result of T_1 concentrating around its mean $\frac{g}{L_{min}}$.
For T_2, we again upper bound the expectation by maximum distance between nodes $\times$ maximum time to spread across an edge. Furthermore, we can upper bound the sum of i.i.d exponential random variables using Chernoff bound to get

$$\mathbb{P}[\sum_{i=1}^{d} Z_i \geq \alpha d] \leq e^{-\psi\alpha d}(1 - \psi)^{-d}$$

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

$$\mathbb{P}[\sum_{i=1}^{d} Z_i > \alpha d] \geq n.2^d e^{-\frac{\alpha d}{2}}$$

Now we use the ccdf definition of expectations

$$\begin{aligned}\mathbb{E}[T_2] &= \int_0^{\infty} \mathbb{P}(T_2 > x)dx \\ &\leq (2\log 2 + 2)d + d\int_{2\log 2+2}^{\infty} \mathbb{P}(T_2 > \alpha d)d\alpha \\ &\leq 3d(n) + 2^d nd \int_{2\log 2+2}^{\infty} e^{-\frac{\alpha d}{2}} d\alpha \\ &= 3d + 2de^{-d}\end{aligned}$$

And since we assumed $d > \log n$, we get the required result.

Theorem 3: (Upper bound, Conductance version). For a graph G and any partition Π(G) = $\bigcup_{i=1}^{g(\Pi)} G_i$ connected by g(Π) subgraphs, each with $s_{min}(\Pi) \leq |G_i| \leq s_{max}(\Pi)$, and conductance $\geq \Psi(\Pi)$:
(a) (mean spreading time)

$$\mathbb{E}[T] \leq k(\Pi).(\log g(\Pi) + 1)$$

where k(Π) = max($\frac{n}{s(\Pi)L_{min}}, \frac{2\log s_{max}(\Pi)}{\Psi(\Pi)}$)
(b) (spreading time concenteration)

$$\mathbb{P}[T \geq \kappa k(\Pi) \log g(\Pi)] \leq \frac{\pi^2}{9\kappa^2(\log g(\Pi))^2}$$

Proof of Theorem 3: Now, T_1 is the same as was in setting of theorem 1. So, we have

$$\mathbb{E}[T_1] \leq \frac{n(\log g + 1)}{s_{min}L_{min}}$$

Also, for variance we have

$$Var[T_1] \leq \frac{n^2}{s_{min}^2 L_{min}^2} \sum_{i=1}^{g} \frac{1}{i^2} \leq \frac{\pi^2 n^2}{6 s_{min}^2 L_{min}^2}$$

For T_2, we have

$$\mathbb{E}[T_2] \leq \frac{2\log s_{max} \log g}{\Psi}$$

and similarly variance can be upper bounded as

$$Var[T_2] \leq 2 \sum_{j=1}^{|V(G_i)|/2} \frac{\pi^2}{6j^2\Psi^2} = \frac{\pi^4}{18\Psi^2}$$

Since we have divided the process into two, using the standard Chernoff ideas would be difficult as computing MGF will be very tedious. Also, the distribution is no longer in standard exponential form, so put bounds on the variance and use **Chebyshev's** inequality, so we have for any $\kappa > 0$:

$$\begin{aligned}&\mathbb{P}[T \geq \kappa k \log g] \leq \mathbb{P}[T_1 + T_2 \geq \kappa k \log g] \\ &\leq \mathbb{P}[T_1 + T_2 \geq \frac{\kappa}{2}\left(\frac{n}{s_{min}L_{min} + \frac{\log s_{max}}{\Psi}}\right) \log g] \\ &\leq \frac{\pi^2(\frac{n^2}{s_{min}^2 L_{min}^2} + \frac{1}{\Psi^2})}{9\kappa^2 \log^2 g(\frac{n}{s_{min}L_min} + \frac{\log s_{max}}{\Psi})^2} \\ &\leq \frac{\pi^2}{9\kappa^2(\log g(\Pi))^2}\end{aligned}$$

as $\log s_{max} > 1$ always.

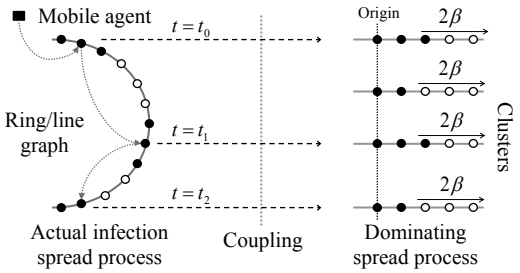

Fig. 1. Dominating the infection spread using independently growing clusters. Here mobile agent represents the external infection agent. Taken from Fig.1 of this paper

IV. Lower Bounds for specific topologies and any policy

Theorem 4: (Lower bound for ring graphs). For a ring graph G_n with n nodes, given $L_{max} \leq 1$. $\forall t \geq 0$, then for any external-spreading policy, we have:

(a) $\mathbb{E}[T_1] \geq \frac{2}{3}\sqrt{n}$

(b) $\mathbb{P}[T < \sqrt{n}/8] \leq 4e^{-\sqrt{n}/8}$

Proof of Theorem 4: The author introduces a new random process $(\tilde{S}(t))_{t\geq 0}$ along with the spreading process $(S^{\mathcal{P}}(t))_{t\geq 0}$ induced by policy $\mathcal{P}$ and has β as the intrinsic spreading rate. The new random process is described as follows:

(a) It has, at all times t, an integer number of points called clusters denoted by $\tilde{C}_t$. $(\tilde{C}_t)_{t\geq 0}$ is a Poisson process with parameter $1(\geq L_{max})$ and at t = 0, $\tilde{C}_0 = 1$ which means there is an initial cluster that has an intrinsic spreading

(b) Each cluster formed adds points to itself which follows a Poisson process of intensity 2β

It can be observed that the number of points $\tilde{S}(t)$(denoted by $\tilde{N}_t$) stochastically dominates $S^{\mathcal{P}}(t)$ for any policy $\mathcal{P}$. This is attributed to the fact that the rate of external infection induced by $\mathcal{P}$ is atmost as that of $\tilde{S}(\cdot)$. Also each cluster in $\tilde{S}(\cdot)$ grows independently and unboundedly without any interference from the neighbouring clusters which is not the case in $S^{\mathcal{P}}(\cdot)$ in which clusters can merge. This can be visualized in Figure 1 where each cluster in the new process grows independently and without interference from each other whereas the in the original process, clusters might merge.

Let $\tilde{T} \triangleq inf\{t \geq 0 : \tilde{N}_t = n\}$ which is the minimum time taken for the number of points in $\tilde{S}(\cdot)$ to reach n. From stochastic dominance, we can say that $\mathcal{N}(S^{\pi}(t)) \leq_{st} \tilde{N}_t$ and therefore for any policy $\mathcal{P}$, we can say:

$$\tilde{T} \leq_{st} T_{\mathcal{P}} \tag{1}$$

Next we calculate $\mathbb{E}[\tilde{N}_t]$

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

Let $\tilde{T}_1, \ldots, \tilde{T}_k$ be the time of creation of k clusters in the event that $\{\tilde{C}_t = k\}$. Then these times are uniformly distributed on the interval [0,t] as $\tilde{C}_t$ is a Poisson process. The expected size of the ith cluster is $2\beta(t - \tilde{T}_i), 1 \leq i \leq k$ as each cluster is growing at a rate of 2β. The 0th cluster initialized at time t = 0 has an expected size of $2\beta t$. Given the creation instants are uniformly distributed, we have $\mathbb{E}[\tilde{T}_i|\tilde{C}_t = k] = t/2$. Thus we have:

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

Now we use **Markov's inequality** as using Chernoff bound is not feasible because MGF of $\tilde{N}_t$ is not trivial to find as the process is not a standard Poisson process. Therefore using Markov's inequality and the CCDF definition of expectations, we get:

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

From (1) , we have for any policy $\mathcal{P}$, and large enough n:

$$\mathbb{E}[T_{\mathcal{P}}] \geq \frac{2}{3\sqrt{\beta}}\sqrt{n}.$$

To prove second part of theorem 4, let $\tilde{X}_i(s)$ denote the size of the ith cluster at time $s \geq \tilde{T}_i$. Therefore we can write:

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

We then apply **Chernoff bound** to $\tilde{C}_t \sim$ Poisson(t) and $\tilde{X}_i(t + \tilde{T}_i) \sim$ Poisson$(2\beta t)$. Chernoff bound applied to a Poisson random variable with parameter λ gives us an upper bound: $\mathbb{P}[Y \geq 2e\lambda] \leq (2e)^{-\lambda}$. This is proved in our assignments and the basic idea is after using chernoff bound, we calculate the MGF of Y and minimize the entire RHS of

the Chernoff bound giving the result. Thus from above we can write:

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

Using the stochastic dominance (1), if $\beta \geq 1$:

$$\mathbb{P}\left[T < \sqrt{\frac{n}{8\beta e^2}}\right] \leq \mathbb{P}\left[\tilde{T} < \sqrt{\frac{n}{8\beta e^2}}\right] = \mathbb{P}\left[\tilde{N}_{\sqrt{\frac{n}{8\beta e^2}}} > n\right] \leq 4e^{-\sqrt{\frac{n}{8\beta e^2}}}.$$

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

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

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

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

In order to prove Theorem 5, we will use the following lemma which uses the theory of first-passage percolation which gives a bound on the extent of infection on an infinite scale at time t

Lemma 1: If $(\tilde{Z}(t))_{t\geq 0}$ represent an infection spread process on the infinite d-dimensional lattice starting from the origin of the lattice at time 0, then $\exists$ positive constants l, c_1, c_2 such that for $t \geq 1$:

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

Proof of Lemma 1: Let us consider

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

which denotes the set of infected nodes at time t in $\tilde{Z}$. *(Theorem 2 in first-passage percolation theory) 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* $t \geq 1$,

$$\mathbb{P}\left[\tilde{B}(t) \subset tB_0\right] \geq 1 - c_1 t^{2d} e^{-c_2\sqrt{t}}. \tag{3}$$

Using (3), for $t \geq 1$, we have:

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

Proof of Theorem 5: We use the concept used in the proof of Theorem 4 and we introduce a new dominating counting process $(\tilde{S}(t))_{t\geq 0}$ which consists of an integer number of clusters denoted by $(\tilde{C}_t)_{t\geq 0}$ which is a Poisson process (with parameter $L_{max}(n)$) and at time t = 0, there is an initial infected node. Also, each cluster grows as per the intrinsic spreading process and independent of each other on an exclusive d-dimensional grid starting from origin. Using similar arguments used in theorem 4, we say that the total number of points in $\tilde{S}(t)$ (denoted by $\tilde{N}_t$) stochastically dominated that in S(t). Let $\tilde{T} \triangleq inf\{t \geq 0 : \tilde{N}_t = n\}$ be the minimum time taken for the number of points in $\tilde{S}(\cdot)$ reaches n. Then we have

$$\mathcal{N}(S(t)) \leq_{st} \tilde{N}_t \ \Rightarrow \ \tilde{T} \leq_{st} T. \tag{4}$$

Let $\tilde{X}_i(s)$ denote the size of the ith cluster created 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 here each $\tilde{X}_i(t+T_i)$ is nothung but the number of infected nodes on an infinite grid following the infection process at time t. Therefore using 2 from Lemma 1 and **Chernoff bound** for $\tilde{C}_t \sim$ Poisson(t$L_{max}(n)$), we have:

$$\begin{aligned}
&\mathbb{P}\left[\tilde{N}_t \geq (2eL_{\max}(n)l^d)t^{d+1}\right] \\
&\leq \mathbb{P}\left[\tilde{C}_t \geq 2eL_{\max}(n)t\right] + \sum_{i=1}^{2eL_{\max}(n)t} \mathbb{P}\left[\tilde{X}_i(t+T_i) \geq t^d l^d\right] \\
&\leq (2e)^{-L_{\max}(n)t} + 2eL_{\max}(n)t \cdot c_1 t^{2d} e^{-c_2\sqrt{t}} \\
&= O(L_{\max}(n)e^{-c_2\sqrt{t}}).
\end{aligned}$$

Now we use the stochastic dominance (4) which gives us:

$$\begin{aligned}
\mathbb{P}\left[T \leq \left(\frac{n}{2eL_{\max}(n)l^d}\right)^{1/(d+1)}\right] &\leq \mathbb{P}\left[\tilde{T} \leq \left(\frac{n}{2eL_{\max}(n)l^d}\right)^{1/(d+1)}\right] \\
&= \mathbb{P}\left[\tilde{N}_{\left(\frac{n}{2eL_{\max}(n)l^d}\right)^{1/(d+1)}} \geq n\right] \\
&= O\left(e^{-c_2\left(\frac{n}{L_{\max}(n)}\right)^{1/(2d+2)}}\right),
\end{aligned} \tag{5}$$

for the appropriate c_2, thus establishing the first part of the theorem. Using 5 along with the fact that $L_{max}(n) = O(n^{1-\epsilon})$ and Borel-Cantelli lemma, we have:

$$\mathbb{P}\left[\tilde{T} \leq \left(\frac{n}{2eL_{\max}(n)l^d}\right)^{1/(d+1)} \text{ for finitely many } n\right] = 1,$$

$$\Rightarrow \liminf_{n\to\infty} \frac{\tilde{T}}{(n/L_{\max}(n))^{1/(d+1)}} \overset{a.s.}{\geq} c_4 \triangleq \frac{1}{(2el^d)^{1/(d+1)}} > 0$$

Using Fatou's lemma, we have:

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

Thus proving $\mathbb{E}[T] \geq \mathbb{E}[\tilde{T}] = \Omega\left((n/L_{\max}(n))^{1/(d+1)}\right)$.

Before moving onto the next Theorem, we would like to introduce a popular family of random graphs known as the

Geometric Random Graph (RGG). It is a random graph model where n points are placed uniformly and iid on a unit square ($[0,1]\times[0,1]$) and an edge exists between two nodes iff the distance between those two nodes is less than r_n, where r_n is often called as the coverage radius.

Theorem 6: (Spreading-time for random external-infection on the RGG) For the planar random geometric graph $G_n(r_n)$, if $r_n \geq \sqrt{\frac{5(1+\gamma)\log n}{n}}$, for random external spreading, we have:

(a) If $\gamma \geq \frac{2}{3}$, then: $\mathbb{E}[T] \leq 2\sqrt[3]{n/L_{\min}}\log n$
(b) For any $\gamma > 0$, choosing $\kappa \geq 1 + 3\gamma/(1+\delta)$ we have:

$$\mathbb{P}[T \geq \kappa\sqrt[3]{n/L_{\min}}\log n] \leq 2n^{-\gamma}$$

Proof of Theorem 6: First the unit square ($[0,1]\times[0,1]$) is divided into square tiles each of length $r_n/\sqrt{5}$. Thus there are a total of $5/r_n^2$ tiles, say $k_1,\ldots,k_{5/r_n^2}$. Now n points are uniformly and randomly thrown onto this unit square. Let $\mathcal{E}$ denote the even that some tile is empty:

$$\begin{aligned}\mathbb{P}[\mathcal{E}] &\leq \frac{5}{r_n^2}\mathbb{P}[\text{tile 1 empty}] = \frac{5}{r_n^2}\left(1-\frac{r_n^2}{5}\right)^n \\ &\leq \frac{n}{\log n}\exp(-\log n) = \frac{1}{\log n} \overset{n\to\infty}{\longrightarrow} 0. \end{aligned} \tag{6}$$

The maximum distance between points in any two adjacent tiles (horizontal or vertical) is exactly r_n (consider the diagonal of the rectangle created) because of the way we have constructed. Therefore any two nodes in adjacent tiles are always connected by an edge, but a node in a tile is not connected to any node in a tile which is atleast three hops away. Let $\triangleq nL_{min}^2$. Now we divide the original square in a different way. We divide it into bigger square chunks each of side length $1/\sqrt[6]{\tilde{n}}$. There are $\sqrt[3]{\tilde{n}}$ square chunks each having a $\frac{\sqrt{5}}{r_n\sqrt[6]{\tilde{n}}}\times\frac{\sqrt{5}}{r_n\sqrt[6]{\tilde{n}}}$ grid of square tiles. In the case when no tile is empty, from the arguments of the previous paragraph, it follows that the diameter D of the subgraph of each square chunk is:

$$D \leq \frac{2\sqrt{5}}{r_n\sqrt[6]{\tilde{n}}} \leq \frac{2}{\sqrt{(1+\gamma)\log n}}\sqrt[3]{\frac{n}{L_{\min}}}.$$

Choosing $n \geq e^{4/1+\gamma}$, an application of Theorem 1 in this case shows that for random external-spreading:

$$\mathbb{E}[T|\mathcal{E}] \leq \sqrt[3]{n/L_{\min}}\cdot(\log n+1).$$

Also note that for any graph, using random external-spreading strategy yields $\mathbb{E}[T] \leq \frac{n}{L_{\min}}$. Thus if $\gamma \geq 2/3$, then combining with (6), we have:

$$\mathbb{E}[T] \leq 2\sqrt[3]{n/L_{\min}}\log n.$$

Now let $\delta \triangleq \frac{2}{3}\log_n\left[\frac{n}{L_{\min}}\right]$, then if $\mathcal{E}$ is given, the each subset in the partition has a size of atleast n^δ. Now, for any $\gamma > 0$ and choosing $\kappa > \left(1+\frac{3\gamma}{1+\delta}\right)$ as per the theorem, using the concentration result from Theorem 1, we get:

$$\mathbb{P}\left[T \geq \kappa\sqrt[3]{n/L_{\min}}\log n|\mathcal{E}\right] \leq \frac{1}{n^\gamma}.$$

Combining this with equation (6), we get:

$$\mathbb{P}[T \geq \kappa\sqrt[3]{n/L_{\min}}\log n] \leq \frac{2}{n^\gamma}.$$

The proof is complete wiht this

Theorem 7: (Lower bound for the RGG) For a planar RGG G_n with coverage radius $r_n = O(\sqrt{log(n)/n})$ having a single node infected at time t = 0 and any spreading policy with $L_{max} = O(n^{1-\epsilon])}$ for some $\epsilon \in (0,1], \exists\beta > 0$ such that:

$$\lim_{n\to\infty}\mathbb{P}\left[T \geq \beta\frac{\sqrt[3]{n/L_{\max}}}{\log^{4/3}n}\right] = 1.$$

Consider an infinite planar grid having one-hop diagonal edges. Let $(S(t))_{t\geq 0}$ denote an infection process with each edge propogating infection at an exponential rate β and correspondingly let I(t) denote the set of infected nodes at time t. Therefore the following lemma helps in giving a bound on the size of I(t).

Lemma 2: There exists c_1 such that for any $c_2 > 0$ and large enough t we have:

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

Proof of lemma 2:

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

Now 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 $||v_{i+1}-v_i||_\infty = 1$. Continuing the above chain of inequalities we have:

$$\begin{aligned}&\sum_{\{v:||v||_\infty=\lfloor ct\rfloor\}}\sum_{v_0,\ldots,v_{\lfloor ct\rfloor}}\mathbb{P}[\exists \text{ a path } r : 0 \to v \text{ passing} \\ &\text{successively through the } v_i, T(r) \leq t] \\ &\leq \sum_{\{v:||v||_\infty=\lfloor ct\rfloor\}}\sum_{v_0,\ldots,v_{\lfloor ct\rfloor}}\mathbb{P}\Bigg[\exists \text{ a path } r \text{ passing} \\ &\text{successively through the } v_i, \sum_{i=0}^{\lfloor ct\rfloor-1}T(v_i,v_{i+1}) \leq t\Bigg] \\ &\leq \sum_{\{v:||v||_\infty=\lfloor ct\rfloor\}}\sum_{v_0,\ldots,v_{\lfloor ct\rfloor}}\mathbb{P}\left[\sum_{i=0}^{\lfloor ct\rfloor-1}T(v_i,v_{i+1}) \leq t\right], \end{aligned} \tag{7}$$

where the second sum runs through all vertices v_i with $v_0 = 0$, $||v_i||_\infty \leq \lfloor ct\rfloor$ and $||v_{i+1}-v_i||_\infty = 1$, and $T(x,y)$ represents the time taken for infection to reach node y from x.

Let $T'(v_i, v_{i+1})$ be random variables identically distributed as $T(v_i, v_{i+1})$ but independent, we will use **Chernoff bound** which gives:

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

Here in the last step we have summed over the vertices and used the fact that infection spread times are translation-invariant, i.e. for any $x, y, a \in \mathbb{Z}^2$,

$$T'(x, y) \overset{d}{=} T(x, y) \overset{d}{=} T(x + a, y + a) \overset{d}{=} T'(x + a, y + a).$$

Consider the neighbor of 0, let it be u ($||u||_\infty = 1$). Then we have $T(0, u) \ge min_{w:||w||_\infty=1} t((0, w))$ where $t(e) \sim \text{Exp}(\mu)$ is the time taken for infection to spread across an edge e of the graph. Number of neighbors of 0 in G is exactly 8 (diagonal neighbors are also considered), therefore $T(0, u)$ stochastically dominates an exponential random variable with parameter 8μ. Let us define $T \sim \text{Exp}(8\mu)$, we have:

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

$$\Rightarrow e^{\psi t}\left(\sum_{\{u:||u||_\infty=1\}} \mathbb{E}\left[e^{-\psi T'(0,u)}\right]\right)^{\lfloor ct\rfloor} \le e^{\psi t}\left(8\left(1 + \frac{\psi}{8\mu}\right)^{-1}\right)^{\lfloor ct\rfloor}. \tag{9}$$

Setting $\psi = 8\mu(8e-1)$ so that $8(1+\psi/8\mu)^{-1} = e^{-1}$, equation 9 becomes:

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

Finally letting $c_1 = 8(8e - 1)$ and $c = c_1\mu + c_2$, we get the desired result from 7 and above:

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

Proof sketch of Theorem 7: The method of proof for theorem 7 is similar to that of theorem 5. We introduce a dominating spreading process that spreads faster than the original infection process and using lemma 2 show that even this fast process must take at least the claimed amount of time to spread. This dominated process $(\tilde{S}(t))_{t\ge0}$ consists of an integer number of clusters denoted by $(\tilde{C}_t)$ with an initial cluster at time 0, the process $(\tilde{C}(t))_{t\ge0}$ is a Poisson process with parameter $1(\ge L_{max}(n))$. Then using this process we use stochastic dominance and then bound it using lemma 2 which gives us 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 atleast $O\left(\frac{\sqrt[3]{n}}{\log^{4/3} n}\right)$ time for spreading to spreading w.h.p.

V. Discussion and Uses

Studying Epidemic spread with external agents also has a twofold utility which is:

(a) During adversarial scenarios such as a malicious epidemics which is a menace as it can spread via both intrinsic and external means, it becomes crucial to study and understand the worst-case infection behavior brought about by external agents so as to take appropriate actions and curb the epidemic.

(b) When looking for cases where optimal and efficient spread of epidemics are required (an optimization problem), we need to find the optimal policy for the external spread so as to disseminate the infection through the network as efficiently as possible. Examples can be viral advertising, network protocol design, diffusion of innovations, and all such scenarios where external agent is useful in spread of desired component

The framework of Epidemic Spreading with external agents can describe a lot of known models for epidemic spreading with long-range interactions. This can be done by specifying $\tilde{L}(t) \in \mathbb{R}_+^{|V|}$ as a function of time t, the network topology and the network-state $S(t)$.

Static Links: Consider a d-dimensional grid with static edges, then the lower bound on spreading time T by setting $L_{max} = L(n)$ in theorem 5 is $\mathbb{E}[T] = \Omega\left(\left(\frac{n}{L(n)}\right)^{\frac{1}{d+1}}\right)$. On combining this with Theorem 2, we can get the same lower bound on the diameter $D(n)$ of the resultant graph which results in $D(n) = \Omega\left((n/L(n))^{\frac{1}{d+1}}\right)$. This has a consequence on small world graphs where 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. This is useful in the sense that it shows it is not possible to obtain such sub-polynomial diameters by adding $O(n^{1-\epsilon})$ edges.

Dynamic Links and Mobile Agents: When considering a grid with additional dynamic links, i.e. long range links that can change their endpoints as time progresses, a more surprising result is obtained. Dynamic links, unlike static links, can be re-used over time to help spread infection. The surprising part is that the lower bound on the spreading time is $\Omega\left(\left(\frac{n}{L}\right)^{\frac{1}{d+1}}\right)$. This is by letting $L = O(n^{1-\epsilon})$ and using Theorem 5.

Sub-Polynomial Spreading Time: So far we have considered examples where the spreading time is polynomial in the graph size. But the techniques discussed in paper do not yield tight bounds in the two extreme regions: where diameter

is sub-polynomial and when the infection rate is high, i.e. $L(t) = \Omega(n)$. There is little work in analysing such regimes and existing work focuses on specific graph and infection models.

Computational Complexity of Fast-Spreading Policies: Another interesting question that the paper has not discussed is the complexity of the design of optimal external-infection policies for general graphs. This is essentially a Markov Decision Process on the space of all subsets of V and is apparently connected to a known NP-complete problem. They do not design such optimal policies but the results do indicate that even simple policies have good approximation ratio. The related problem is to minimize the maximum distance of any node from k-seed nodes which is chosen from V. This is a special case of the *k-center* problem which is shown to be a NP-hard problem. The problem of finding a good partition is similar to the k-centre problem except the fact the we do not know k. We can still use the k-center algorithms by following the steps. We use the 2-approximate algorithm for k-center problem and execute it sequentially from $\{1, 2, 4, \ldots, 2^{logn}\}$. We stop when the maximum diameter of the resulting partition is less than the current value of k. As the diameter decreases with k, it can be easily shown that the resulting partition is a 4-approximation to the problem.

VI. Conclusions

The paper considers the spread of epidemics with the SI model and assisted by external agents that can infect nodes devoid of any infected neighbours. Upper bounds on spreading times are obtained by partitioning the main graph into subgraphs and dividing the spreading process into a slower combination of two separate processes, one with only external agents infecting across subgraphs and one with only intrinsic spreading infecting nodes within its own subgraph only. This helped to give upper bounds in terms of the graph partition properties when the spreading policy is known. Further, lower bounds on the spreading time are derived by upper bounding the number of nodes that can be infected within a fixed time, by creating independent copies of the graph at every node that is externally infected. This works in settings where the spreading policy can be anything but the graph topology is known and mathematically tractable, such as rings, grids and random geometric graphs. These results can be used and applied in several epidemic phenomena, and some possible ones are discussed in the context of static and dynamic links and order-wise spreading times for various graphs. Random external spreading policy is also shown to be order-wise optimal (upto a log n factor) for minimizing the spreading time.

References

Siddhartha Banerjee et al., "Epidemic Spreading With External Agents," journal-article, IEEE TRANSACTIONS ON INFORMATION THEORY, vol. 60, July 2014.

"On The Speed of Convergence in First-Passage Percolation on JSTOR," Www.Jstor.Org, n.d., https://www.jstor.org/stable/2959634.