Epidemics with SI Dynamics and External Agents (New Results and Attempts)

Tamojeet Roychowdhury, Joel Anto Paul, Anway Deshpande

***Abstract*—The original paper looks at epidemics with external agents that can spread the infection even when the nodes are not directly connected, under the SI (susceptible-infected) dynamics. They derive upper bounds on the spreading time in both concentration and expectation, for general graphs with specific external infection policies. For this, an exponential distribution of spreading time, both intrinsic from node-to-node and external, is assumed. We re-derive these bounds in similar forms to the original paper assuming a Rayleigh distribution which has sub-Gaussian tail. Further, the original paper derives lower bounds on the spreading times for specific graph structures and general external infection policies. We utilize these (in their original form, with no change to the exponential time distribution in this case) to define vaccination conditions which are sufficient to make the proportion (with high probability) of infected nodes among total nodes go to zero as the total number of nodes asymptotically approaches infinity.**

I. Introduction

The original paper looks at infections spreading via SI dynamics across connected nodes in a graph $G(V, E), |V| = n$ (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.

By the memorylessness of the exponential distribution, treating the infection times as being exponential from the last infection is identical to treating the infection times as being exponential from the start of the epidemic. This simplifies many of the bounds and derivations that the paper follows.

Often, however, in real-life diseases, once a person is infected, their infectivity often first rises with time till a peak, and then decreases rapidly down to nearly zero. Thus instead of the exponential, where the infectivity peaks at $t = 0$, we wish to consider a different distribution, where the infectivity peaks somewhere later than 0 and then decays again. We considered the lognormal, Gamma and Rayleigh distributions as possible candidates for the same. We rejected the lognormal and Gamma distributions because they had heavy tails (super-exponential), and most of the derivations rely on tail bounds.

tamojeet@iitb.ac.in

For the first section, the Rayleigh distribution was finally chosen because of its sub-Gaussian tail; our results for the first section are derived for sub-Gaussian tails in general, while the original paper addresses sub-exponential tails. This part spins off from the upper bounds on specific policies for general graphs in the paper.

The Problem of Non-Memorylessness: Exponential spreading times are widely used in several classes of these epidemic dynamics and modelling, owing to its easy translation into a continuous-time Markov chain (CTMC) with the sojourn times being equivalent to the node-to-node infection time, and the states being the number of infected nodes. This CTMC form is especially useful and hence extensively used in the susceptible-infected-susceptible (SIS) model, where the number of infected states can both increase and reduce. Another advantage of exponentials is that of a constant rate irrespective of when which node was infected, which is the property of memorylessness unique to exponential (and in discrete, geometric) distributions. This obviously fails when we shift to the Rayleigh distribution of spreading times.

However, for the SI model, since the number of infected nodes only increases with time, the CTMC approach is not needed at all. Further, the first two proofs we formulate with the Rayleigh distribution use either the *sum* of many such i.i.d. distributions, or the *maximum* of the same. These bounds do not need the property of memorylessness. In the third proof (conductance version), the probability of infection not spreading across any of the $j\Psi$ edges in some time a is considered. This time a is measured from the last infection, not from the time the concerned edge acquired one infected node. But here the bound still works because then the actual probability we want is that of $\mathbb{P}(T > b + a)$ where $(b + a)$ is the total time since that edge's end-node acquired infection, and necessarily larger than a as no node is infected in the last a duration. Hence, $\mathbb{P}(T > b + a) \leq \mathbb{P}(T > a)$, so our derived bounds still work.

The second section, spun off from lower bounds on specific graphs for any policies, looks at vaccination strategies for these specific graphs. Ideally, we would want as small as possible a portion of the population to get infected. We consider the case where the number of people infected is $o(n)$ w.h.p, thus as n increases, the proportion of infected people $o(n)/n$ goes to 0. Using the lower bounds on spreading time given in the paper, we derive sufficient conditions on the vaccination rate to contain the infection to $o(n)$.

II. New Upper Bounds for Specific Policies

Our spreading times follow the given distributions:

(a) The times for external spreading have a Rayleigh distribution, with start times at $t = 0$, and parameter $\sigma_{e,i}$ for i^{th} node. The start time at $t = 0$ is based on the external factor infecting nodes randomly irrespective of the last infection time.
(b) The time for intrinsic infection to spread from one node to its neighbour via an edge has a Rayleigh distribution for each edge, with parameter σ_0. Here, of course, the start time for an edge is the time when one of its endpoints gets infected, and not $t = 0$.

Lemma 1: We note that the Rayleigh distribution has tails satisfying $\mathbb{P}(X > a) = \exp\left(-\frac{a^2}{2\sigma^2}\right)$. Then if we use the bounds on the moment-generating function for a distribution with Gaussian tails, as done in lemma 1.3-1.4 in this note by MIT OCW,

$$\mathbb{E}[e^{sX}] \leq e^{4\sigma^2 s^2}$$

Following this with the derivation on the bound of expectation of Z = the maximum of n i.i.d. random variables with Gaussian tails in this answer from math.stackexchange, we get

$$\mathbb{E}[Z] \leq 2\sigma\sqrt{2\log n}$$

We outline the proof as it is used extensively used later. *Proof of Lemma 1*: We have by definition $\mathbb{P}(X > a) = \exp\left(-\frac{a^2}{2\sigma^2}\right)$. Then, since X is nonnegative

$$\begin{aligned}
\mathbb{E}[|X|^k] &= \int_0^\infty \mathbb{P}(X^k > a)da \\
&= \int_0^\infty \mathbb{P}(X > a^{1/k})da \\
&= \int_0^\infty \exp\left(-\frac{a^{2/k}}{2\sigma^2}\right)da \\
&= (2\sigma^2)^{k/2}k\int_0^\infty e^u u^{k/2-1}du \\
&= (2\sigma^2)^{k/2}k\Gamma(k/2)
\end{aligned}$$

since the last integral is the definition of the gamma function.

For bounding the MGF, we use the Taylor series expansion of the exponential

$$\begin{aligned}
\mathbb{E}[e^{sX}] &\leq 1 + \sum_{k=2}^\infty \frac{s^k\mathbb{E}[|X|^k]}{k!} \\
&\leq 1 + \sum_{k=2}^\infty \frac{(2\sigma^2 s^2)^{k/2}k\Gamma(k/2)}{k!} \\
&= 1 + \sum_{k=1}^\infty \frac{(2\sigma^2 s^2)^{k/2}(2k)\Gamma(k)}{(2k)!} \\
&\quad + \sum_{k=1}^\infty \frac{(2\sigma^2 s^2)^{k/2}(2k+1)\Gamma(k+1/2)}{(2k+1)!}
\end{aligned}$$

Note that both $\Gamma(k)$ and $\Gamma(k+1/2)$ are bounded by $k!$, and that $(2k)! > 2(k!)^2$. Using these relations, we get

$$\begin{aligned}
\mathbb{E}[e^{sX}] &\leq 1 + \left(1 + \sqrt{\frac{\sigma^2 s^2}{2}}\right)\sum_{k=1}^\infty \frac{(2\sigma^2 s^2)^k k!}{(2k)!} \\
&= e^{2\sigma^2 s^2} + \sqrt{\frac{2\sigma^2 s^2}{2}}(e^{2\sigma^2 s^2} - 1) \\
&\leq e^{4\sigma^2 s^2}
\end{aligned}$$

Finally, we use the MGF bound to derive a bound on Z, the maximum of n i.i.d. random variables satisfying the above condition. By Jensen's inequality,

$$e^{s\mathbb{E}[Z]} \leq \mathbb{E}[e^{sZ}] \leq \mathbb{E}[\max_i e^{sX_i}]$$

By union bound,

$$e^{s\mathbb{E}[Z]} \leq n\mathbb{E}[e^{sX_i}] \leq ne^{4\sigma^2 s^2}$$

Therefore,

$$\mathbb{E}[Z] \leq \frac{\log n}{s} + 4\sigma^2 s$$

Minimizing over s gives

$$\mathbb{E}[Z] \leq 2\sigma\sqrt{2\log n}$$

We now give the graph-based theorems. Let Π denote a partition of the graph G into $g(\Pi)$ subgraphs such that subgraph has

(a) Diameter $\leq d(\Pi)$
(b) Number of nodes $\geq s(\Pi)$
(c) Conductance $\geq \Psi(\Pi)$

We drop the Π in parentheses in further proofs for brevity. Each theorem we find is immediately followed by the proof, with most proofs proceeding on similar lines of stochastic dominance as used in the paper. We say a random variable A stochastically dominates B if $\mathbb{P}(A > a) \geq \mathbb{P}(B > a) \ \ \forall a \in \mathbb{R}$

Theorem 1: For random external spreading, where $\sigma_{e,i} = \sigma_e$, the total spreading time T satisfies the following bounds:
(a)

$$\mathbb{E}[T] \leq \sqrt{\frac{\pi}{2s}}\sigma_e + 2\sigma_0 d\sqrt{2\log n}$$

(b)

$$\mathbb{P}(T \geq a) \leq \exp\left(-\frac{sa^2}{2\sigma_e^2}\right) + d\exp\left(-\frac{a^2}{2\sigma_0^2}\right)$$

Proof of Theorem 1: Similar to the original paper, consider a slower spreading process with two times, T_1 and T_2

- 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.

It is clear that $T_1 + T_2$ stochastically dominates T, since we are limiting the agent of spread in each phase when dividing up T_1 and T_2. A formal proof is given in the paper. We separate bounds on T_1 and T_2, making extensive use of the fact that

the Rayleigh distribution has (sub-)Gaussian tail, so its ccdf $\mathbb{P}(A > a) = e^{-a^2/2\sigma^2}$.

Bounds on T_1: Note that the event of $T_1 > a$ implies there exists at least one subgraph with all nodes untouched by the external agent. Since the size of each subgraph is at least s, this means that there is at least one set of s points not infected by the external agent.

$\{T_1 > a\} \implies \{$at least one subgraph of s nodes has all of its external infection times $> a\}$ hence

$$\begin{aligned}\mathbb{P}(T_1 > a) &\leq \mathbb{P}(T_{\text{external}} > a)^s \\ &\leq \left[\exp\left(-\frac{a^2}{2\sigma_e^2}\right)\right]^s \\ &\leq \exp\left(-\frac{sa^2}{2\sigma_e^2}\right)\end{aligned}$$

For bounding the expectation, we use the standard result for a nonnegative continuous random variable A, $\mathbb{E}[A] = \int_0^\infty \mathbb{P}(A > a)da$. Thus for the above case, the final expression for $\mathbb{P}(T_1 > a)$ integrates out to the integral of the right half of a Gaussian, but without the appropriate scaling. This gives us the bound of

$$\mathbb{E}[T_1] \leq \sqrt{\frac{\pi}{2s}}\sigma_e$$

.

Bounds on T_2: Here we follow the same method as done in the original paper. Note that after a node in every subgraph is infected, then by the maximum diameter condition, the longest direct path from the first infected node to any other node is d.

Then T_2 can be bounded by the product of the maximum time taken for the infection to spread across *one* edge out of n edges, multiplied by the maximum number of edges it needs to traverse on a direct path from the first infected node to any other node, equal to d. As done in the paper:

$$T_2 \leq \sum_{i=1}^{d(\Pi)} \max_j \hat{T}_j$$

where $\hat{T}_j$ is the time taken to spread from j^{th} node. This is given by the maximum of n i.i.d. Rayleigh distributions. If Z is the maximum of n i.i.d. random variables with (sub-)Gaussian tails, then Z satisfies the following bound as given in Lemma 1:

$$\mathbb{E}[Z] \leq 2\sigma\sqrt{2\log n}$$

Hence our expectation bound becomes

$$\mathbb{E}[T_2] \leq 2d\sigma_0\sqrt{2\log n}$$

Note that this expression is exactly the same as the one in the original paper, with the only difference being that the bound of $\log n$ for sub-exponential tails has now been tightened to $\sqrt{\log n}$ for sub-Gaussian tails, which is expected.

We also want a concentration/tail bound on T_2. We note that if T_2 is more than a, that means at least one node in a subgraph did not receive the infection via a path no longer than d. Mathematically,

$\{T_2 > a\} \implies \{\exists$ a path of $\leq d$ edges, which had intrinsic spreading time $> a\} \implies \{$at least one edge of d edges had spreading time $> a$ $\}$

Thus, by using a union bound, we have

$$\begin{aligned}\mathbb{P}(T_2 > a) &\leq \mathbb{P}(\text{one of } d \text{ edges has spread time} > a) \\ &\leq d\exp\left(-\frac{a^2}{2\sigma_0^2}\right)\end{aligned}$$

Finally, we substitute this in the two inequalities below, which come from stochastic dominance, to get the final expressions:

$$\begin{aligned}\mathbb{E}[T] &\leq \mathbb{E}[T_1] + \mathbb{E}[T_2] \\ \mathbb{P}(T > a) &\leq \mathbb{P}(T_1 > a) + \mathbb{P}(T_2 > a)\end{aligned}$$

Theorem 2: For a greedy subgraph policy, where $\sigma_{e,i} = \sigma_e$ for one particular node in each subgraph (so as to make spreading to all subgraphs fastest), the bounds from Theorem 1 are slightly modified into

(a)

$$\mathbb{E}[T] \leq \sqrt{\frac{\pi}{2(n-g)}}\sigma_e + 2\sigma_0 d\sqrt{2\log n}$$

(b)

$$\mathbb{P}(T \geq a) \leq \exp\left(-\frac{(n-g)a^2}{2\sigma_e^2}\right) + d\exp\left(-\frac{a^2}{2\sigma_0^2}\right)$$

Proof of Theorem 2: The bounds on T_2 remain the same. For T_1, the only difference for the greedy subgraph policy is that now one subgraph not getting infected means that at most only one node each from the remaining $(g-1)$ subgraphs can have been infected, and hence all the remaining $n - g + 1 > n - g$ nodes are un-infected. Thus,

$$\begin{aligned}\mathbb{P}(T_1 > a) &\leq \mathbb{P}(T_{\text{external}} > a)^{(n-g)} \\ &\leq \left[\exp\left(-\frac{a^2}{2\sigma_e^2}\right)\right]^{(n-g)} \\ &\leq \exp\left(-\frac{(n-g)a^2}{2\sigma_e^2}\right)\end{aligned}$$

The remaining derivation remains exactly the same.

Theorem 3: In this case, we also need the *maximum* size (number of nodes) of a subgraph, call it s_m. For random external spreading, where $\sigma_{e,i} = \sigma_e$, we can also bound the total spreading time in terms of the minimum conductance of each subgraph Ψ as follows

(a)

$$\mathbb{E}[T] \leq \sqrt{\frac{\pi}{2s}}\sigma_e + \frac{2\sigma_0\sqrt{2\log g}\log s_m}{\Psi}$$

(b)

$$\mathbb{P}(T \geq a) \leq \exp\left(-\frac{sa^2}{2\sigma_e^2}\right) + 2gh\left(\frac{1-h^{s_m/2}}{1-h}\right)$$

where $h = \exp\left(-\frac{\Psi a^2}{2s_m^2\sigma_0^2}\right)$

Proof of Theorem 3: The bounds for T_1 remain the same as for Theorem 1 since it is the same scenario of external spreading across subgraphs.

For T_2, consider the transition time $\hat{T}_{j,j+1}$ between the time the j^{th} node gets infected to the time when $(j+1)^{th}$ node gets infected. Since the conductance of the graph is Ψ, any set of j infected nodes has at least $\Psi \times \min(j, |V_k| - j)$ edges

connecting them to the non-infected nodes, where $|V_k|$ is the number of vertices in the k^{th} subgraph. We first consider the case where j is less than $|V_k|/2$.

Therefore there are at least $j\Psi$ paths that can transmit the infection to the $(j+1)^{th}$ node. Also note that the number of transition times $\hat{T}_{j,j+1}$ that need to elapse before the entire subgraph is infected is at most s_m. Consequently, if the total spread time is more than a, and given maximum number of spread times is s_m, at least one spread time should have been more than a/s_m We have two event dependencies as follows

$\{T_2 > a\} \implies \{\hat{T}_{j,j+1} > a/s_m$ for at least one j out of s_m edges needed to be covered$\} \implies \{$In time interval a/s_m, no edge among $j\Psi$ edges available has transmitted infection$\}$

Note that the probability of the last event occurring is $\exp\left(-j\Psi\frac{a^2}{2s_m^2\sigma_0^2}\right) = h^j$. For the case of $j > |V_k|/2$, we can use the same arguments replacing j with $|V_k| - j$ as was used in the paper.

Using a union bound over all available j for one subgraph and over all g subgraphs,

$$\begin{aligned}\mathbb{P}(T_2 > a) &\leq 2g\sum_{j=1}^{s_m/2}\exp\left(-j\Psi\frac{a^2}{2s_m^2\sigma_0^2}\right)\\ &= 2g\sum_{j=1}^{s_m/2}h^j = 2gh\left(\frac{1-h^{s_m/2}}{1-h}\right)\end{aligned}$$

For the bound on expectation, we have that T_2 is the maximum of the intrinsic spreading time from first node to last node, maximized over all graphs (indexed by k). Mathematically

$$T_2 = \max_k\sum_{j=1}^{s_m}\hat{T}_{j,j+1} \leq \sum_{j=1}^{s_m}\left(\max_k\hat{T}_{j,j+1}\right)$$

The term in brackets is a sum of at most g i.i.d. Rayleigh distributions, with $\sigma^2 = \sigma_0^2/j\Psi$. The expectation of this, according to Lemma 1, will satisfy

$$\mathbb{E}[T_2] \leq \sum_{j=1}^{s_m}\left(2\frac{\sigma_0}{j\Psi}\sqrt{2\log g}\right) \leq \frac{2\sigma_0\sqrt{2\log g}\log s_m}{\Psi}$$

III. Original Lower Bounds and Vaccination Rates

We use the derived lower bounds in the paper on the spreading time to find sufficient vaccination rates to ensure the proportion of infected nodes going to 0 as n grows to ∞ with high probability (henceforth called *contained infection*).

We use the fact that the spreading time for a fixed number of nodes being lower bound, corresponds to the number of nodes infected in a fixed time being upper bound.

Note that for these, we do not re-derive from scratch but instead continue with the original paper's assumption of both the external infection times and intrinsic spreading times being exponentially distributed.

Theorem 4: For rings/linear graphs of number of nodes n, a vaccination rate of order $\Theta(t^\alpha)$ for any $\alpha > 2$ is sufficient to ensure that with high probability, the ratio of infected people to the total population is $o(n)$. Further, a vaccination rate of $\Theta(t^\alpha)$ for any $\alpha > 1$ is necessary to ensure the same condition.

Proof of Theorem 4: The necessity condition is easy to prove. Consider only the external infection spread process M_t, with inter-infection times distributed exponentially with some rate λ. Clearly, the number of nodes infected by the full process N_t is more than the number of nodes infected only by the external infection process. Since it is a Poisson process, $\mathbb{E}[N_t] > \mathbb{E}[M_t] = \lambda t$. Suppose vaccination happens at a $\Theta(n)$ rate of rt.

Then in expectation, $\mathbb{E}[N_t] + rt = n$ or $E[M_t] + rt \leq n$, which means $\mathbb{E}[N_t]$ is at least $\Theta(n)$ in order, contrary to what we want. This shows that vaccination rate should be of order $> t^1$ for contained infection.

For the sufficiency condition, we use the same method as used in the original paper for the proof. If S_t is the original infection process under vaccination, we construct a stochastically dominating process $\tilde{S}_t$ which consists of two simultaneous spreads

- The external spread infects nodes irrespective of their position in the ring. At time t, let the number be C_t. Each of these are called clusters
- Each node starts an individual line of infection by intrinsic spreading $N_{i,t}$ for i^{th} cluster, at some rate 2β where β is the rate of original intrinsic spread. Let the total number of these separately infected nodes be N_t.

$\tilde{S}_t$ dominates the original spreading $S(t)$ because the original infection also consists of the same two spreads, but in the intrinsic spreading the total infected number cannot freely keep increasing - at some point it hits another infected node or a vaccinated node along the ring and stops.

Further, the number of nodes infected in a given time under vaccination, $\tilde{S}_t$, is less than the number of nodes infected without vaccination in the same time, call it $\hat{S}_t$. The bounds derived in the paper are then on $\hat{S}_t$, infection under no vaccination, which we can exploit for bounds on $\tilde{S}_t$ and S_t.

Note that the maximum time that can elapse before the graph hits steady state (all nodes either infected or vaccinated) can be $t = (n/r)^{1/\alpha}$, since the rate of vaccination is $n(t) = rt^\alpha$.

Hence, we bound the probability that the number of nodes infected in time $t = (n/r)^{1/\alpha}$, given by $S_{(n/r)^{1/\alpha}}$.

$$\mathbb{P}(S_{(n/r)^{1/\alpha}} > a) \leq \mathbb{P}(\hat{S}_{(n/r)^{1/\alpha}} > a)$$

The term on RHS is exactly the term bounded in the original paper. As given in the original paper,

$$\mathbb{P}(\hat{S}_t > 8\beta e^2t^2) \leq (2e)^{-t} + 2et\cdot(2e)^{-2\beta t}$$

Substituting $t = (n/r)^{1/\alpha}$,

$$\mathbb{P}\left(\hat{S}_{(n/r)^{1/\alpha}} > 8\beta e^2\left(\frac{n}{r}\right)^{2/\alpha}\right) \leq \Theta(e^{-n^{2/\alpha}})$$

The bound on RHS goes to 0, as n becomes very large. Then this shows that for large n, the number of infected nodes is atmost of order $O\left(n^{2/\alpha}\right)$ with high probability. Substituting $\alpha > 2$ completes the proof.

Theorem 5: For d-dimensional grids, a vaccination rate of $\Theta(t^\alpha)$ with $\alpha > d+1$ is sufficient to ensure the constrained infection condition.

Proof of Theorem 5: We use the same technique as for Theorem 4. The maximum time before hitting steady state is again $t = (n/r)^{1/\alpha}$. As per the bound for the unvaccinated infection derived in the paper,

$$\mathbb{P}\left(\tilde{N}_t > (2eL_{max})l^d t^{d+1}\right) \leq O\left(L_{max}e^{-\sqrt{t}}\right)$$

Substituting $t = (n/r)^{1/\alpha}$

$$\mathbb{P}\left(\tilde{N}_t > (2eL_{max})l^d (n/r)^{(d+1)/\alpha}\right) \leq O\left(L_{max}e^{-n^{(d+1)/2\alpha}}\right)$$

Again the RHS bound goes to 0 or large n, and substituting $\alpha > d+1$ gives us the order of infected nodes proportion to be at most $o(n)$.

We could not find a similar useful corollary for theorem 6. Our next theorem is therefore a corollary of theorem 7 in the paper.

Theorem 6: For planar geometric random graph with coverage radius $r = O(\sqrt{\log n/n})$, and any spreading policy with L_{max} of order less than n, a vaccination rate of $\Theta(t^\alpha)$ with $\alpha > 3$ is sufficient to ensure the constrained infection condition.

Proof of Theorem 6: We again use the same technique. The proof sketch for Theorem 7 in the paper says that in time t, for unvaccinated population, the number of clusters is at most $O(t)$ with high probability and the number of infected nodes in each cluster is at most $O(t^2 \log^4 n)$. The total number of infected nodes is then $O(t^3 \log^4 n)$. If the maximum time before hitting steady state is $(n/r)^{1/\alpha}$, the total infected nodes is then $O(n^{3/\alpha} \log^4 n)$. Substituting $\alpha > 3$ gives us the required constrained infection condition.