

isid/ms/2012/03

February 24, 2012

<http://www.isid.ac.in/~statmath/eprints>

On the Expected Total Number of Infections for Virus Spread on a Finite Network

ANTAR BANDYOPADHYAY

FARKHONDEH SAJADI

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi-110 016, India

On the Expected Total Number of Infections for Virus Spread on a Finite Network

Antar Bandyopadhyay*
Farkhondeh Sajadi†

Theoretical Statistics and Mathematics Unit,
Indian Statistical Institute, Delhi Centre,
7 S. J. S. Sansanwal Marg,
New Delhi 110016
INDIA

February 23, 2012

Abstract

In this paper we consider a simple virus infection spread model on a finite population of n agents connected by some neighborhood structure. Given a graph G on n vertices, we begin with some fixed number of initial infected vertices. At each discrete time step, an infected vertex tries to infect its neighbors with probability $\beta \in (0, 1)$ independently of others and then it dies out. The process continues till all infected vertices die out. We focus on obtaining proper lower bounds on the expected number of ever infected vertices. We obtain a simple lower bound, using *breadth-first search* algorithm and show that for a large class of graphs which can be classified as the ones which locally “look like” a tree in sense of the *local weak convergence* [1], this lower bound gives better approximation than some of the known approximations through matrix-method based upper bounds [3].

AMS 2000 Subject Classifications: Primary: 60K35, 05C80; secondary: 60J85, 90B15

Keywords and phrases: Bread-first search, local weak convergence, random r -regular graphs, susceptible infected removed model, virus spread.

*E-Mail: antar@isid.ac.in

†E-Mail: sajadi7r@isid.ac.in

1 Introduction

1.1 Background and Motivation

Often it is observed that the normal operation of a system which is organized in a network of individual machines or agents is threatened by the propagation of a harmful entity through the network. Such harmful entities are often termed as a *viruses*. For example the Internet, as a network is threatened by the computer viruses and worms which are self-replicating pieces of code, that propagate in a network of computers. These codes use a number of different methods to propagate, for example, an e-mail virus typically sends copies of itself to all addresses in the address book of the infected machine. Weaver et. al. [7] gives a good survey of different techniques of propagation for computer viruses.

In this paper we use a simple *susceptible infected removed (SIR)* model which was studied by Draief, Ganesh and Massoulié in 2008 [3]. In this model, each susceptible agent, can be infected by its infected neighbors at a rate, proportional to their number and remains infected till it is removed after an unit time. While it is infected, it has the potential to infect its neighbors. In general, removal can correspond to a quarantining of a machine from the network or patching the machine. In this model, it is assumed that once a node is removed, it is “out of the network”. That is, it can no longer be susceptible or infected. Such a model is justified, provided the epidemic spread happens at a much faster rate than the rate of patching of the susceptible machines.

The study of mathematical models for epidemic spread has a long history in biological epidemiology and in the study of computer viruses. One of the first work in this area was by Kermack and Mckendrick [5], where they established the first stochastic theory for epidemic spread. They also proved the existence of an epidemic threshold, which determines whether the epidemic will spread or die out. As mentioned in [3], earlier work mainly focused on finding or approximating the *law of large numbers* limit where the stochastic behavior was approximated by its mean behavior and hence mainly studied deterministic models. More recent works [2, 6], have focused on stochastic nature of the models and have tried to prove asymptotic distribution of the number of survivors, using a key concept called *basic reproductive number* R_0 , which is defined as the expected number of secondary infective, caused by a single primary infective. This concept of basic reproductive number is well defined under the *uniform mixing* assumption, that is, when any infective can infect any susceptible equally likely, and hence the underlying network is given by a complete graph. For a general network, where basic reproductive number may become vertex dependent, it is not clear how to use this concept effectively. As in [3], in this work we would like to study this model on a general network.

1.2 Model

We consider a closed population of n agents, connected by a network structure, given by an undirected graph $G = (V, E)$ with vertex set V , containing all the agents and edge set E . A vertex can be in either of the three states, namely, *susceptible (S)*, *infected (I)* or *removed (R)*. At the beginning, the initial set of infected is assumed to be non-empty and all others are susceptible. The evolution of the epidemic is described by the following discrete time model:

- After an unit epoch of time, each infected vertex instantaneously tries to infect each susceptible

neighbor with probability $\beta \in (0, 1)$ independent of all others.

- Each infected vertex is removed from the network after an unit time.

Mathematically, at an integer multiple of unit time, say t , if a susceptible vertex v has $I_v(t)$ neighbors who are infected, then the probability of v being infected instantaneously is $1 - (1 - \beta)^{I_v(t)}$ and each susceptible vertex gets infected independently. Also an infected vertex remains in the network only for an unit time, after that it tries to infect its susceptible neighbors and then it is immediately removed.

As pointed out by [3], this is a simple model, falling in the class of models known as Reed-Forest Models, where infection period is deterministic and is same for every vertex. It is worth noting that the evolution of the epidemic can be modeled as a Markov chain.

It is interesting to note here that, the model is essentially same to the i.i.d. Bernoulli bond percolation model with parameter β [10]. This is because the set of ever infected (or removed) vertices is same as the union of connected open components of i.i.d. bond percolation on G , containing all the initial infected vertices. Although for percolation, it is customary to work with an infinite graph G . If G is the complete graph K_n , then this model is fairly well studied in literature which is known as the *binomial random graph*, also known as Erdős-Rényi random graph [4, 11].

Like in [3], our goal is to study the total number of vertices that eventually become infected (and hence removed) without specifying the underlying network. In [3], the authors derived explicit upper bounds to the expected number of vertices ever infected which depends on both the size of the network as well as the infection rate β . These bounds also needed assumption of “small” value for β . Unfortunately, the work [3] did not provide any indication of whether the derived upper bounds are good approximations to the exact quantity of interest. In this work we derive a simple lower bound to the expected number of vertices ever infected which works for every infection rate $0 < \beta < 1$. Our lower bound is based on the *breadth-first search (BFS)* algorithm and hence easily computable for any general finite network G . We also prove that, under certain assumptions on the qualitative behavior of the underlying graph, namely if G “locally looks like a tree” in the sense of Aldous and Steele [1] *local weak convergence*, then our lower bound is asymptotically exact for “small” β , thus provides a good approximation when the network is “large”. As we will see later, for such graphs G , the range we cover for β always include the range in which the upper bounds from [3] works and in all these cases, the upper bound over estimates the exact quantity.

1.3 Outline

In the following section, we state and prove our main results. Section 3 gives several examples where our lower bounds works and gives asymptotically correct answer. Finally in Section 4 we summarize the merits of work and indicate some of its limitations as well.

2 Main Results and Proofs

We will denote by $Y^{G,I}$, the total number of vertices ever infected when the epidemic runs on a network G and the infection starts at the vertices in $I \subseteq V$. Note that $Y^{G,I}$ implicitly depends on the size of the network. In Subsection 2.1 we present the results, when the epidemic starts with only one

infected vertex. We generalize these results for epidemic starting with more than one infection, which are presented in Subsection 2.2. In both cases, our results rely on a specific search algorithm, known as *breadth-first search (BFS)*. We briefly describe the algorithm here.

Step-0 Input graph G with a linear ordering of its vertices say $V := \{v_0, v_1, v_2, \dots, v_{n-1}\}$.

Let $T \leftarrow \{v_0\}$ and $N \leftarrow \{v_0\}$.

Step-1 Write $N = \{v_{i_1}, v_{i_2}, \dots, v_{i_r}\}$ for some $r \geq 1$ such that $i_1 < i_2 < \dots < i_r$.

Step-2 For $l = 1$ to r find all neighbors u of v_{i_l} which are not in T , put

$$N' \leftarrow N' \cup \left\{ u \mid u \sim v_{i_l} \text{ and } u \notin T \right\}$$

and update T as

$$T \leftarrow T \cup \left\{ u \mid u \sim v_{i_l} \text{ and } u \notin T \right\}.$$

Step-3 Update $N \leftarrow N'$.

Step-4 Go to **Step-1** unless vertex set of T is same as that of V .

Step-5 Stop with output T as the BFS spanning tree with root v_0 .

Note that the BFS spanning tree is not necessarily unique, it depends on the starting point v_0 which is typically called the root and also it depends on the ordering of the vertices in which the exploration of neighbors is done in **Step-2**. Also note that iff G is a tree to start with then, BFS spanning tree is just itself.

2.1 Starting with Only One Infected Vertex

Our first result gives a lower bound to the expected total number of vertices ever infected starting with exactly one infected vertex.

Theorem 1 *Let G be an arbitrary finite graph and $v_0 \in V$ be a fixed vertex of it. Let T be a spanning tree of the connected component of G containing the vertex v_0 and rooted at v_0 . Let $Y^{T, \{v_0\}}$ be the total number of vertices ever infected when the epidemic runs only on T and starting with exactly one infection at v_0 . Then*

$$\mathbf{E} \left[Y^{T, \{v_0\}} \right] \leq \mathbf{E} \left[Y^{G, \{v_0\}} \right] \text{ for all } 0 < \beta < 1. \quad (1)$$

Moreover, if T is a BFS spanning tree of the connected component of v_0 rooted at v_0 , then

$$\mathbf{E} \left[Y^{T, \{v_0\}} \right] \leq \mathbf{E} \left[Y^{\mathcal{T}, \{v_0\}} \right] \leq \mathbf{E} \left[Y^{G, \{v_0\}} \right] \text{ for all } 0 < \beta < 1. \quad (2)$$

Proof: First we observe that if $H \subseteq G$, is a spanning sub-graph of G and v_0 is a vertex in both H and G , then by definition $Y^{H, \{v_0\}} \leq Y^{G, \{v_0\}}$. This proves (1).

For the second part, we note that if T is a spanning tree of G with root v_0 , then $d_G(v, v_0) \leq d_T(v, v_0)$ for all $v \in V$, where d_G and d_T are the graph distance functions on G and T respectively. Moreover, the BFS algorithm preserves the distances, so if \mathcal{T} is a BFS spanning tree with root $\{v_0\}$ then we must

have $d_G(v, v_0) = d_{\mathcal{T}}(v, v_0)$ for all $v \in V$. Thus $d_{\mathcal{T}}(v, v_0) \leq d_T(v, v_0)$ for all $v \in V$. Now from the model description, it follows that for any spanning tree T with root v_0 we have

$$\mathbf{E} \left[Y^{T, \{v_0\}} \right] = \sum_{v \in V} \beta^{d_{\mathcal{T}}(v, v_0)}.$$

So we conclude that

$$\mathbf{E} \left[Y^{T, \{v_0\}} \right] = \sum_{v \in V} \beta^{d_{\mathcal{T}}(v, v_0)} \leq \sum_{v \in V} \beta^{d_T(v, v_0)} = \mathbf{E} \left[Y^{T, \{v_0\}} \right],$$

as $0 < \beta < 1$. ■

Let $\text{LB}^{G, \{v_0\}} := \mathbf{E} \left[Y^{T, \{v_0\}} \right]$ be the lower bound obtained through BFS algorithm for a BFS spanning tree \mathcal{T} of G , rooted at v_0 . Then from the proof of Theorem 1 we get that

$$\text{LB}^{G, \{v_0\}} = \sum_{v \in V} \beta^{d_G(v, v_0)}, \quad (3)$$

which is free of the choice of the BFS spanning tree. Later, we will see that, this helps us to generalize the lower bound for epidemic starting with more than one infected vertices at the beginning. We also note that $\text{LB}^{G, \{v_0\}}$ can easily be computed using the breadth-first search algorithm described before.

Our next result shows that if we have a “large” finite graph G on n vertices and the epidemic starts with exactly one infected vertex v_0 , such that any cycle containing v_0 is “relatively large”, that is of order $O(\log n)$, then the lower bound $\text{LB}^{G, \{v_0\}}$ given above, is asymptotically same as the exact quantity $\mathbf{E} \left[Y^{G, \{v_0\}} \right]$.

To state the result rigorously, we use the following graph theoretic notations. Given a graph G , a fixed vertex v_0 of G and $d \geq 1$, let $V_d(G)$ be the set of vertices of G which are at a *graph distance* at most d from v_0 in G . Let $N_d(G, v_0)$ be the induced sub-graph of G on the vertices $V_d(G)$.

Theorem 2 *Let $\{(G_n, v_0^n)\}_{n \geq 1}$ be a sequence of rooted graphs with roots $\{v_0^n\}_{n \geq 1}$ such that there exists a sequence $\alpha_n = O(\log n)$ with $N_{\alpha_n}(G_n, v_0^n)$ is a tree for all $n \geq 1$. Then, there exists $0 < \beta_0 \leq 1$, such that for all $0 < \beta < \beta_0$*

$$\frac{\mathbf{E} \left[Y^{G_n, \{v_0^n\}} \right]}{\text{LB}^{G_n, \{v_0^n\}}} \longrightarrow 1 \text{ as } n \rightarrow \infty. \quad (4)$$

Proof: Let \mathcal{T}_n be a BFS spanning tree rooted at v_0^n of the graph G_n and as defined before let $\text{LB}^{G_n, \{v_0^n\}} = \mathbf{E} \left[Y^{\mathcal{T}_n, \{v_0^n\}} \right]$. Then

$$\begin{aligned} \text{LB}^{G_n, \{v_0^n\}} &\leq \mathbf{E} \left[Y^{G_n, \{v_0^n\}} \right] \\ &\leq \mathbf{E} \left[Y^{N_{\alpha_n}(G_n, v_0^n), \{v_0^n\}} \right] + \mathbf{E} \left[Y^{N_{\alpha_n}(G_n, v_0^n), \{v_0^n\}} \right] \times \beta^{\alpha_n} \times n \\ &\leq \text{LB}^{G_n, \{v_0^n\}} + \text{LB}^{G_n, \{v_0^n\}} \times \beta^{\alpha_n} \times n, \end{aligned} \quad (5)$$

where the last inequality follows from the fact that $N_{\alpha_n}(G_n, v_0^n)$ is a tree and hence is a subtree of \mathcal{T}_n . This proves (4) since by assumption $\alpha_n = O(\log n)$. ■

Although the assumption in the above theorem, may seem to be too restrictive but it is satisfied by many examples including the n -cycle (see Subsection 3.2). The method of the proof on the other hand, helps us to generalize the result for a large class of graphs including certain random graphs.

Following Aldous and Steele [1], we say a sequence of rooted random or deterministic graphs $\{(G_n, v_0^n)\}_{n \geq 1}$ with roots $\{v_0^n\}_{n \geq 1}$ converges to a random or deterministic graph (G_∞, v_0^∞) in the sense of *local weak convergence (l.w.c)* and write $(G_n, v_0^n) \xrightarrow{l.w.c.} (G_\infty, v_0^\infty)$ if for any $d \geq 1$,

$$\mathbf{P}(N_d(G_n, v_0^n) \cong N_d(G_\infty, v_0^\infty)) \longrightarrow 1 \text{ as } n \rightarrow \infty. \quad (6)$$

Note that for a deterministic sequence graphs, (6) really means that the event occurs for “large” enough n .

Theorem 3 *Let $\{(G_n, v_0^n)\}_{n \geq 1}$ be a sequence of rooted deterministic or random graphs with deterministic or randomly chosen roots $\{v_0^n\}_{n \geq 1}$. Suppose that for each G_n the maximum degrees of a vertex is bounded by Δ . Suppose there is a rooted deterministic or random tree \mathcal{T} with root ϕ such that*

$$(G_n, v_0^n) \xrightarrow{l.w.c.} (\mathcal{T}, \phi) \text{ as } n \rightarrow \infty. \quad (7)$$

Let $LB^{G_n, \{v_0\}} := \mathbf{E} \left[Y^{\mathcal{T}_n, \{v_0^n\}} \right]$ where \mathcal{T}_n is a BFS spanning tree rooted at v_0^n of the graph G_n .

Then for $\beta < \frac{1}{\Delta}$

$$\left(\mathbf{E} \left[Y^{G_n, \{v_0^n\}} \right] - LB^{G_n, \{v_0\}} \right) \longrightarrow 0 \text{ as } n \rightarrow \infty. \quad (8)$$

Moreover for $\beta < \frac{1}{\Delta}$ we also get

$$\lim_{n \rightarrow \infty} LB^{G_n, \{v_0\}} = \lim_{n \rightarrow \infty} \mathbf{E} \left[Y^{G_n, \{v_0^n\}} \right] = \mathbf{E} \left[Y^{\mathcal{T}, \phi} \right]. \quad (9)$$

Proof: Let \mathcal{T}_n be a BFS spanning tree rooted at v_0^n of the graph G_n and as earlier let $LB^{G_n, \{v_0\}} = \mathbf{E} \left[Y^{\mathcal{T}_n, \{v_0^n\}} \right]$. Let E_n be the event $[N_d(G_n, v_0^n) \cong N_d(\mathcal{T}, \phi)]$. Then from Theorem 1

$$LB^{G_n, \{v_0\}} \leq \mathbf{E} \left[Y^{G_n, \{v_0^n\}} \right] = \mathbf{E} \left[Y^{G_n, \{v_0^n\}} \mathbf{1}_{E_n} \right] + \mathbf{E} \left[Y^{G_n, \{v_0^n\}} \mathbf{1}_{E_n^c} \right]. \quad (10)$$

Now under our assumption, degree of any vertex of G_n is bounded by Δ and $\beta < \frac{1}{\Delta}$, so using Theorem 2.3 of [3] we get

$$\mathbf{E} \left[Y^{G_n, \{v_0^n\}} \mathbf{1}_{E_n^c} \right] \leq \frac{1}{1 - \beta \Delta} \mathbf{P}(E_n^c). \quad (11)$$

Further note that $N_d(G_n, v_0^n)$ is a tree rooted at v_0^n if E_n occurs and thus on E_n , $N_d(G_n, v_0^n)$ is a sub-tree of \mathcal{T}_n . So

$$Y^{N_d(G_n, v_0^n), \{v_0^n\}} \mathbf{1}_{E_n} \leq Y^{N_d(\mathcal{T}_n, v_0^n), \{v_0^n\}} \mathbf{1}_{E_n} \leq Y^{\mathcal{T}_n, \{v_0^n\}} \mathbf{1}_{E_n}.$$

Thus we get

$$\begin{aligned} \mathbf{E} \left[Y^{G_n, \{v_0^n\}} \mathbf{1}_{E_n} \right] &\leq \mathbf{E} \left[Y^{N_d(\mathcal{T}_n, v_0^n), \{v_0^n\}} \mathbf{1}_{E_n} \right] + \beta^d \mathbf{E} \left[Y^{G_n, \partial_d^* N_d(G_n, v_0^n)} \right] \\ &\leq \mathbf{E} \left[Y^{N_d(\mathcal{T}_n, v_0^n), \{v_0^n\}} \mathbf{1}_{E_n} \right] + \beta^d \frac{1}{1 - \beta\Delta} \mathbf{E} \left[|\partial_d^* N_d(G_n, v_0^n)| \right] \end{aligned} \quad (12)$$

$$\begin{aligned} &\leq \text{LB}^{G_n, \{v_0\}} + \beta^d \frac{1}{1 - \beta\Delta} \mathbf{E} \left[Y^{G_n, \{v_0^n\}} \right] \\ &\leq \text{LB}^{G_n, \{v_0\}} + \beta^d \frac{1}{(1 - \beta\Delta)^2}, \end{aligned} \quad (13)$$

where $\partial_d^* N_d(G_n, v_0^n)$ denotes the infected vertices in G_n after d units of time starting with one infection at vertex v_0^n . In the last two inequalities we use Theorem 2.3 of [3].

So finally combining (10), (13) and (11) we get that for $\beta < \frac{1}{\Delta}$ and for any $d \geq 1$ we have

$$\left(\mathbf{E} \left[Y^{G_n, \{v_0^n\}} \right] - \text{LB}^{G_n, \{v_0\}} \right) \leq \beta^d \frac{1}{(1 - \beta\Delta)^2} + \frac{1}{1 - \beta\Delta} \mathbf{P}(E_n^c). \quad (14)$$

Now under assumption (7), we have $\lim_{n \rightarrow \infty} \mathbf{P}(E_n^c) = 0$ so we conclude that for any $d \geq 1$

$$\limsup_{n \rightarrow \infty} \left(\mathbf{E} \left[Y^{G_n, \{v_0^n\}} \right] - \text{LB}^{G_n, \{v_0\}} \right) \leq \beta^d \frac{1}{(1 - \beta\Delta)^2}. \quad (15)$$

This proves (8) by taking $d \rightarrow \infty$ as $\beta < 1$.

Now for proving (9) we first note that from definition

$$\mathbf{E} \left[Y^{N_d(G_n, v_0^n), \{v_0^n\}} \mathbf{1}_{E_n} \right] = \mathbf{E} \left[Y^{N_d(\mathcal{J}, \phi), \{\phi\}} \right] \mathbf{P}(E_n). \quad (16)$$

Thus arguing similar to the derivation of the equation (14) we get

$$\left| \mathbf{E} \left[Y^{G_n, \{v_0^n\}} \right] - \mathbf{E} \left[Y^{N_d(\mathcal{J}, \phi), \{\phi\}} \mathbf{P}(E_n) \right] \right| \leq \beta^d \frac{1}{(1 - \beta\Delta)^2} + \frac{1}{1 - \beta\Delta} \mathbf{P}(E_n^c). \quad (17)$$

Now from (7) we conclude that the degree of any vertex of \mathcal{J} is also bounded by Δ . So using Theorem 2.3 of [3] we get that for $\beta < \frac{1}{\Delta}$

$$\mathbf{E} \left[Y^{N_d(\mathcal{J}, \phi), \{\phi\}} \right] < \frac{1}{1 - \beta\Delta}.$$

Moreover from definition, $Y^{N_d(\mathcal{J}, \phi), \{\phi\}} \uparrow Y^{\mathcal{J}, \phi}$ as $d \rightarrow \infty$. So by MCT we get

$$\mathbf{E} \left[Y^{\mathcal{J}, \phi} \right] < \frac{1}{1 - \beta\Delta}. \quad (18)$$

Now under our assumption (7) we have $\mathbf{P}(E_n) \rightarrow 1$. So from (17) we conclude that

$$\lim_{n \rightarrow \infty} \mathbf{E} \left[Y^{G_n, \{v_0^n\}} \right] = \mathbf{E} \left[Y^{\mathcal{J}, \phi} \right]. \quad (19)$$

Thus using (8), it follows that

$$\lim_{n \rightarrow \infty} \text{LB}^{G_n, \{v_0\}} = \lim_{n \rightarrow \infty} \mathbf{E} \left[Y^{G_n, \{v_0^n\}} \right] = \mathbf{E} \left[Y^{\mathcal{J}, \phi} \right].$$

This completes the proof. ■

Following is immediate but interesting application of the above theorem which gives an explicit formula for the limit of epidemic spread on a randomly selected r -regular graph when infection starts from an randomly chosen vertex.

Theorem 4 *Suppose G_n is a graph selected uniformly at random from the set of all r -regular graphs on n vertices where we assume nr is an even number. Let v_0^n be an uniformly selected vertex of G_n . Then for $\beta < \frac{1}{r}$*

$$\lim_{n \rightarrow \infty} \mathbf{E} \left[Y^{G_n, \{v_0^n\}} \right] = \frac{1 + \beta}{1 - (r - 1)\beta}. \quad (20)$$

We note that in this case, the upper bound given in [3] is $\frac{1}{1-r\beta}$ when $\beta < \frac{1}{r}$ which is strictly bigger than the exact answer given in (20).

Proof: It is known [4, 1] that if G_n is a graph selected uniformly at random from the set of all r -regular graphs on n vertices, where nr is even and v_0^n be a randomly selected vertex of G_n then

$$(G_n, v_0^n) \xrightarrow{l.w.c.} (\mathbb{T}_r, \phi), \quad (21)$$

where \mathbb{T}_r is the infinite r -regular tree with root say ϕ . The result then follows from Theorem 3 and equation (38). ■

2.2 Starting with More than One Infected Vertices

Now suppose instead of one infection, we start with k infected vertices given by $I := \{v_{0,1}, v_{0,2}, \dots, v_{0,k}\}$. Following theorem gives a lower bound similar to that of Theorem 1.

Theorem 5 *Let G be an arbitrary finite graph and $I := \{v_{0,j}\}_{j=1}^k$ be a fixed set of k vertices. Let \mathcal{T} be a spanning forest of the connected components of G containing the vertices in I with exactly k trees which are rooted at the vertices in I . Then*

$$\mathbf{E} [Y^{\mathcal{T}, I}] \leq \mathbf{E} [Y^{G, I}] \quad \text{for all } 0 < \beta < 1. \quad (22)$$

Moreover, if \mathcal{T} is a breath-first-search spanning forest of the connected components of G containing the vertices in I with exactly k trees which are rooted at the vertices in I then

$$\mathbf{E} [Y^{\mathcal{T}, I}] \leq \mathbf{E} [Y^{\mathcal{T}, I}] \leq \mathbf{E} [Y^{G, I}] \quad \text{for all } 0 < \beta < 1. \quad (23)$$

Given a finite labeled graph G and a fixed set of vertices $I = \{v_{0,j}\}_{j=1}^k$ of it, by a *breath-first-search spanning forest* of the connected components of G containing the vertices in I with exactly k trees which are rooted at the vertices in I , we mean a spanning forest of G with exactly k connected components which are rooted at the vertices $\{v_{0,1}, v_{0,2}, \dots, v_{0,k}\}$, that are obtained through the *breath-first-search* algorithm, starting at some vertex $v \in I$ and assuming that all the vertices $\{v_{0,1}, v_{0,2}, \dots, v_{0,k}\}$ are at the same level. Alternatively, we can consider a new graph G^* which is same as G except it has one “artificial” vertex, say v^* which is connected to the vertices $v_{0,1}, v_{0,2}, \dots, v_{0,k}$ through k “artificial”

edges and we perform the BFS algorithm on G^* starting with the vertex v^* obtaining a BFS spanning tree, say \mathcal{T}^* of G^* rooted at v^* . Then a *breath-first-search spanning forest* of G with exactly k trees which are rooted at the vertices $\{v_{0,1}, v_{0,2}, \dots, v_{0,k}\}$ is given by the forest $\mathcal{T}^* \setminus \{v^*\}$. This alternative description is quite useful in practice. Note that if $\{\mathcal{T}_i\}_{1 \leq i \leq k}$ be the k connected components, rooted respectively at $\{v_{0,1}, v_{0,2}, \dots, v_{0,k}\}$ of \mathcal{T} , then from definition, the following identity holds for every $\beta \in (0, 1)$

$$\mathbf{E}[Y^{\mathcal{T}, I}] = \sum_{i=1}^k \mathbf{E}[Y^{\mathcal{T}_i, I}] = \frac{\mathbf{E}[Y^{\mathcal{T}^*, \{v^*\}}] - 1}{\beta}. \quad (24)$$

Using the above identity, we can now generalize all the results of the previous section for epidemic spread starting with more than one infected vertices.

We write $\text{LB}^{G, I}$ for $\mathbf{E}[Y^{\mathcal{T}, I}]$ which is the lower bound for starting with k infected vertices given by I . Observe that from equation (24) one can represent

$$\text{LB}^{G, I} = \sum_{i=1}^k \mathbf{E}[Y^{\mathcal{T}_i, I}], \quad (25)$$

where $\mathcal{T} = \bigcup_{i=1}^k \mathcal{T}_i$ is as above. It is worth nothing here that the lower bound $\text{LB}^{G, I}$ does not depend on the choice of \mathcal{T} but the representation given in equation (25) uses a specific choice of \mathcal{T} .

Theorem 6 *Let $\{(G_n, I_n)\}_{n \geq 1}$ be a sequence of graphs where each G_n has k -roots given by $I_n := \{v_{0,1}^n, v_{0,2}^n, \dots, v_{0,k}^n\}$ such that there exists a sequence $\alpha_n = O(\log n)$ with $N_{\alpha_n}(G_n, I_n) := \bigcup_{j=1}^k N_{\alpha_n}(G_n, v_{0,j}^n)$ is a forest with k components. Then there exists $0 < \beta_0 \leq 1$, such that for all $0 < \beta < \beta_0$*

$$\frac{\mathbf{E}[Y^{G_n, I_n}]}{\text{LB}^{G_n, I_n}} \longrightarrow 1 \text{ as } n \rightarrow \infty. \quad (26)$$

The proof of this result is exactly similar to that of Theorem 2 when one uses the identity (24). The details are thus omitted.

Our next result is parallel to the Theorem 3 which needs a generalization of the concept of local weak convergence which was introduced by Wästlund [8].

We will say a sequence of random or deterministic graphs $\{G_n\}_{n \geq 1}$ with k roots given by $I_n := \{v_{0,1}^n, v_{0,2}^n, \dots, v_{0,k}^n\}$, $n \geq 1$ converges to a random or deterministic graph G_∞ with k -roots say $I_\infty := \{v_{0,1}^\infty, v_{0,2}^\infty, \dots, v_{0,k}^\infty\}$ in the sense of *local weak convergence (l.w.c)* and write $(G_n, I_n) \xrightarrow{\text{l.w.c.}} (G_\infty, I_\infty)$ if for any $d \geq 1$

$$\mathbf{P}(N_d(G_n, v_{0,j}^n) \cong N_d(G_\infty, v_{0,j}^\infty) \text{ for all } 1 \leq j \leq k) \longrightarrow 1 \text{ as } n \rightarrow \infty. \quad (27)$$

Note that for a deterministic sequence graphs, (27) really means that the event occurs for “large” enough n .

Theorem 7 *Let $(G_n)_{n \geq 1}$ be a sequence of deterministic or random graphs. Suppose each G_n has deterministic or randomly chosen k roots given by $I_n := \{v_{0,1}^n, v_{0,2}^n, \dots, v_{0,k}^n\}$ and maximum degree*

of each G_n is bounded by Δ . Suppose $\mathcal{T} := \bigcup_{j=1}^k \mathcal{T}_j$ is a forest with k rooted trees with roots $I_\infty := \{\phi_1, \phi_2, \dots, \phi_k\}$. We assume that

$$(G_n, I_n) \xrightarrow{l.w.c.} (\mathcal{T}, I_\infty) \text{ as } n \rightarrow \infty. \quad (28)$$

Then for $\beta < \frac{1}{\Delta}$

$$(\mathbf{E} [Y^{G_n, I_n}] - LB^{G_n, I_n}) \longrightarrow 0, \quad (29)$$

as $n \rightarrow \infty$. Moreover

$$\lim_{n \rightarrow \infty} LB^{G_n, I_n} = \lim_{n \rightarrow \infty} \mathbf{E} [Y^{G_n, I_n}] = \mathbf{E} [Y^{\mathcal{T}, I_\infty}] = \sum_{j=1}^k \mathbf{E} [Y^{\mathcal{T}_j, \{\phi_j\}}]. \quad (30)$$

Proof: For each $n \geq 1$ as done above we define a new rooted graph G_n^* with artificial vertex v_n^* which is connected to the k roots in I_n of G_n through k artificial edges. Also we consider \mathcal{T}^* defined similarly with an artificial root ϕ^* connecting to $\{\phi_1, \phi_2, \dots, \phi_k\}$. Then our assumption of local weak convergence (28) is equivalent to

$$(G_n^*, v_n^*) \xrightarrow{l.w.c.} (\mathcal{T}^*, \phi^*). \quad (31)$$

This together with the relation (24) and Theorem 3 completes the proof. \blacksquare

It is worth noting that in case $\{\mathcal{T}_j\}_{1 \leq j \leq k}$ are i.i.d. (if they are random) or isomorphic (if they are constant) then equation (30) can be reformulated as

$$\lim_{n \rightarrow \infty} LB^{G_n, I_n} = \lim_{n \rightarrow \infty} \mathbf{E} [Y^{G_n, I_n}] = \mathbf{E} [Y^{\mathcal{T}, I_\infty}] = k \mathbf{E} [Y^{\mathcal{T}_1, \{\phi_1\}}]. \quad (32)$$

Like in the case of starting with one infection, following is an immediate application of the above results.

Theorem 8 *Suppose G_n is a graph selected uniformly at random from the set of all r -regular graphs on n vertices where we assume nr is an even number. Let $I_n := \{v_{0,1}^n, v_{0,2}^n, \dots, v_{0,k}^n\}$ be k uniformly and independently selected vertices of G_n . Then for $\beta < \frac{1}{r}$*

$$\lim_{n \rightarrow \infty} \mathbf{E} [Y^{G_n, I_n}] = k \frac{1 + \beta}{1 - (r-1)\beta}. \quad (33)$$

Proof: Since the vertices in I_n are selected uniformly at random so using [1] we get

$$(G_n, I_n) \xrightarrow{l.w.c.} (\mathcal{T}_r, I_\infty), \quad (34)$$

where $I_\infty := \{\phi_1, \phi_2, \dots, \phi_k\}$ and \mathcal{T}_r is a forest with k infinite r -regular trees with roots in I_∞ . The result then follows from Theorems 7 and 4. \blacksquare

Once again we note that in this case, the upper bound $\frac{k}{1-r\beta}$ given in [3] for $\beta < \frac{1}{r}$, is strictly bigger than the exact answer given in (33) and the gap increases with k , the initial number of infections.

3 Examples

3.1 Tree

If G is a tree and the epidemic starts with only one infection at a vertex say ϕ which we call the root, then from the construction of the lower bound it is clear that $\text{LB}^{G, \{\phi\}} = \mathbf{E}[Y^{G, \{\phi\}}]$. In certain cases one can find explicit formula for this quantity. Two such examples are discussed below.

Regular Tree Consider a rooted r -array tree ($r \geq 2$), with height m , denote it by $T(r, m)$. In $T(r, m)$ every internal vertex except the root ϕ has degree r . A vertex v is said to be an internal vertex if it has a neighbor which is not on the unique path from v to ϕ . We assume that the degree of the root ϕ is $(r - 1)$. Let $\mu_m := \mathbf{E}[Y^{T(r, m), \{\phi\}}]$. Note that the total number of vertices in $T(r, m)$ is $\frac{r^{m+1}-1}{r-1}$. Now, to calculate the exact value of μ_m we note that

$$\mu_m = 1 + (r - 1)\beta\mu_{m-1} \tag{35}$$

which gives the formula

$$\mu_m = \frac{[(r - 1)\beta]^{m+1} - 1}{(r - 1)\beta - 1}. \tag{36}$$

As $T(r, m)$ is a tree so the lower bound is exact, that is, $\text{LB}^{T(r, m), \{\phi\}} = \mu_m$. Now upper bound from [3] is $\frac{1}{1-r\beta}$ which only works for $\beta < \frac{1}{r}$. If $\beta < \frac{1}{r}$ then by Theorem 3 we get

$$\mathbf{E}[Y^{T(r), \{\phi\}}] = \lim_{m \rightarrow \infty} \mu_m = \frac{1}{1 - (r - 1)\beta}, \tag{37}$$

where $T(r)$ is the rooted infinite r -regular tree, where each vertex except the root ϕ has degree r and the root degree is $(r - 1)$.

We observe a gap between the lower bound (which in this case the exact answer) to that of the upper bound from [3].

Now let \mathbb{T}_r be the infinite r -regular tree where each vertex including the root has degree r . Such a tree can be viewed as disjoint union of r rooted infinite r -regular trees whose roots are joint to the root, say ϕ of \mathbb{T}_r . Thus from (37) we get that for $\beta < \frac{1}{r}$

$$\text{LB}^{\mathbb{T}_r, \{\phi\}} = \mathbf{E}[Y^{\mathbb{T}_r, \{\phi\}}] = 1 + \frac{r\beta}{1 - (r - 1)\beta} = \frac{1 + \beta}{1 - (r - 1)\beta}. \tag{38}$$

Galton-Watson Tree Consider a Galton-Watson branching process starting with one individual. Let the mean of the offspring distribution be $c > 0$. We denote the random tree generated by this process as $\text{GW}(c)$ with root ϕ . Once again, as discussed above since $\text{GW}(c)$ is a tree, so $\text{LB}^{\text{GW}(c), \{\phi\}} = \mathbf{E}[Y^{\text{GW}(c), \{\phi\}}]$. Now in this case, the epidemic process starting with only one infection at ϕ , results to a Galton-Watson branching process starting with one individual as the root and with mean of the new progeny distribution as βc . So in particular if $\beta < \frac{1}{c}$ then from standard branching process theory $\mathbf{E}[Y^{\text{GW}(c), \{\phi\}}] < \infty$ and is equal to $\frac{1}{1-\beta c}$ [9].

3.2 Cycle

Cycle graph is a graph that consists of a single cycle. We denote the cycle with n vertices by C_n . For simplicity we assume n is odd and then from the BFS algorithm, it is immediate that starting with one infected individual, say at v_0^n , we have

$$\text{LB}^{C_n, \{v_0\}} = 1 + 2 \left(\beta + \beta^2 + \cdots + \beta^{\frac{n-1}{2}} \right) \quad (39)$$

which converges to $\frac{1+\beta}{1-\beta}$ as $n \rightarrow \infty$ for any $0 < \beta < 1$. Now it is clear from the definition that

$$(C_n, v_0^n) \xrightarrow{l.w.c.} (\mathbb{Z}, 0) . \quad (40)$$

Thus using Theorem 3 we conclude that if $\beta < \frac{1}{2}$ then

$$\lim_{n \rightarrow \infty} \text{LB}^{C_n, \{v_0^n\}} = \lim_{n \rightarrow \infty} \mathbf{E} \left[Y^{C_n, \{v_0^n\}} \right] = \frac{1 + \beta}{1 - \beta} . \quad (41)$$

But, in fact this holds for any $0 < \beta < 1$. This because for cycle, the assumption in Theorem 2 holds for $\alpha_n = n/3$ (say). Thus using the proof of Theorem 2 we conclude that the equation (41) holds for any $0 < \beta < 1$.

Now if the epidemic starts with k initial infected vertices given by $I_n := \{v_{0,1}^n, v_{0,2}^n, \dots, v_{0,k}^n\}$ which are uniformly distributed, then it is easy to see that

$$(C_n, I_n) \xrightarrow{l.w.c.} (\mathbb{Z}_j, 0)_{1 \leq j \leq k} , \quad (42)$$

where \mathbb{Z}_j is just a copy of \mathbb{Z} . Then by Theorem 7 we conclude that for $0 < \beta < \frac{1}{2}$,

$$\lim_{n \rightarrow \infty} \text{LB}^{C_n, I_n} = \lim_{n \rightarrow \infty} \mathbf{E} \left[Y^{C_n, I_n} \right] = k \frac{1 + \beta}{1 - \beta} . \quad (43)$$

But in fact, because of exactly similar reason given above, we can use Theorem 6 with $\alpha_n = O(n)$ to conclude that (43) holds for all $0 < \beta < 1$.

3.3 Generalized Cycle

Suppose in a cycle graph we choose randomly without replacement $2m$ vertices and connect these vertices by joining edges between them where $m \geq 1$ is fixed. We call this graph a Generalized Cycle and denote it by $\text{GC}(n, m)$. Now consider the epidemic model on this graph with one initial infected site v_0^n . For large enough n , the probability of having at least one of the m pairs inside a neighborhood of v_0^n with radius r is given by

$$1 - \left(1 - \frac{2r(2r+1)}{n(n-1)} \right)^m$$

which tends to zero as $n \rightarrow \infty$. Therefore, a fixed neighborhood of the root is a tree with high probability, in fact it is isomorphic to a neighborhood of integer line. Hence by Theorem 3 it follows that for $\beta < \frac{1}{2}$

$$\lim_{n \rightarrow \infty} \text{LB}^{\text{GC}(n, m), \{v_0^n\}} = \lim_{n \rightarrow \infty} \mathbf{E} \left[Y^{\text{GC}(n, m), \{v_0^n\}} \right] = \frac{1 + \beta}{1 - \beta} . \quad (44)$$

Similarly if we start with k initial infected sites, say $I_n := \{v_{0,j}^n\}_{j=1}^k$ which are chosen uniformly at random, then it is easy to see that

$$(\text{GC}(n, m), I_n) \xrightarrow{l.w.c.} (\mathbb{Z}_j, 0)_{1 \leq j \leq k}, \quad (45)$$

where \mathbb{Z}_j is just a copy of \mathbb{Z} . Thus by Theorem 7 we get

$$\lim_{n \rightarrow \infty} \text{LB}^{\text{GC}(n, m), I_n} = \lim_{n \rightarrow \infty} \mathbf{E} \left[Y^{\text{GC}(n, m), I_n} \right] = k \frac{1 + \beta}{1 - \beta}, \quad (46)$$

when $\beta < \frac{1}{3}$, because the maximum degree in $\text{GC}(n, m)$ is 3.

3.4 Cube graph

Cube graph is the graph obtained from the vertices and edges of the 3-dimensional unit cube. We denote it by Q_3 . Suppose initially only the vertex $(0, 0, 0)$ is infected. Consider a BFS spanning tree \mathcal{T} of Q_3 rooted at $(0, 0, 0)$. Since Q_3 has only 8 vertices so $Y^{\mathcal{T}, \{(0,0,0)\}}$ takes values $\{0, 1, 2, 3, 4, 5, 6, 7\}$ and

$$\begin{aligned} \text{LB}^{\mathcal{T}, \{(0,0,0)\}} &= \mathbf{E} \left[Y^{\mathcal{T}, \{(0,0,0)\}} \right] \\ &= 1 + \beta(1 - \beta)^2[1 + (1 - \beta) + (1 - \beta)^2] \\ &\quad + 4\beta^2[(1 - \beta)^2 + (1 - \beta)^3 + (1 - \beta)^4] \\ &\quad + 3\beta^3[(1 - \beta) + (1 - \beta)^2 + 6(1 - \beta)^3 + (1 - \beta)^4] \\ &\quad + 4\beta^4[6(1 - \beta)^2 + 4(1 - \beta)^3] \\ &\quad + 5\beta^5[2(1 - \beta) + 6(1 - \beta)^2] \\ &\quad + 24\beta^6(1 - \beta) \\ &\quad + 7\beta^7 \\ &= 1 + 3\beta + 3\beta^2 + \beta^3 \\ &= (1 + \beta)^3. \end{aligned}$$

In general, the d -dimensional cube graph say Q_d is a d -regular graph which has $n = 2^d$ vertices. Following a similar calculation as done above, one can show that for an epidemic starting at one vertex, the lower bound obtained in Theorem 1 for the expected total number of vertices ever infected is given by $(1 + \beta)^d$.

In this example computation of the exact value of $\mathbf{E} \left[Y^{Q_d, \{(0,0,0)\}} \right]$ is difficult, but we note that there is a gap between the upper bound obtained in [3], namely $\frac{1}{1-d\beta}$ which is valid only when $\beta < \frac{1}{d}$ and our lower bound. Although this is an example which does not fall under any of the theorem we discuss in this paper and hence we are not sure if the lower bound gives better approximation.

4 Discussion

The goal of this study has been to get a better idea of the expected total number of vertices ever infected with as little assumption as possible on the underlying graph G . Our approach has been to find an appropriate lower bound to this expectation. Although from practical point of view, approximation

from above with an upper bound is a more conservative method. As shown in the examples given in Section 3, the only known upper bounds obtained in [3] often over estimate the exact quantity. Moreover the upper bounds in [3] are defined only for “small” values of the parameter β . For an arbitrary finite network, we have obtained a lower bound for the expectation of ever infected vertices for any value of the parameter β which is computable through the breadth-first search algorithm. Theorems 2, 3, 6 and 7 show that this lower bound is asymptotically exact for a large class of graphs when β value is “small”, which always includes the region on which the upper bounds from [3] are defined.

However, we would also like to mention here that even though the lower bound we present, works for any infection parameter $0 < \beta < 1$ but if the underlying graphs has many loops, such as the complete graph K_n , then it does not necessarily give a good approximation. To see this, consider the complete graph K_n and suppose that the epidemic starts at a fixed vertex v_0 . Then the lower bound $\text{LB}^{K_n, \{v_0\}} = 1 + (n - 1)\beta$. Now, let X_1 be the number of infected vertices at time $t = 1$. In this case it is easy to see that $X_1 \sim \text{Binomial}(n - 1, \beta)$. Let u be one of $n - 1 - X_1$ vertices which are not infected at time $t = 1$. Since K_n is the complete graph, so the conditional probability of u becomes infected at time $t = 2$ given X_1 is $1 - (1 - \beta)^{X_1}$. Hence

$$\begin{aligned} \mathbf{E} \left[Y^{K_n, \{v_0\}} \right] &\geq 1 + (n - 1)\beta + \mathbf{E} \left[(n - 1 - X_1) \left(1 - (1 - \beta)^{X_1} \right) \right] \\ &= 1 + (n - 1)\beta + (n - 1) - (n - 1) (1 - \beta^2)^{n-1} \\ &\quad - (n - 1)\beta + (n - 1)\beta(1 - \beta) (1 - \beta^2)^{n-2} \end{aligned}$$

Therefore we get

$$\limsup_{n \rightarrow \infty} \frac{\mathbf{E} \left[Y^{K_n, \{v_0\}} \right] - \text{LB}^{K_n, \{v_0\}}}{\text{LB}^{K_n, \{v_0\}}} \geq \frac{1 - \beta}{\beta}. \quad (47)$$

where $\text{LB}^{K_n, \{v_0\}} := \mathbf{E} \left[Y^{T_n, \{v_0\}} \right]$.

Here, it is worth mentioning that for the complete graph if we start with one infected vertex, then as discussed in Section 1 the set of ever infected vertices is no other than an Erdős-Rényi random graph with parameter n and β . Thus asymptotic behavior of $\mathbf{E} \left[Y^{K_n, \{v_0\}} \right]$ is well understood in the literature [4, 11].

References

- [1] Aldous, D., J. and Steele M., J., (2004), The objective method: probabilistic combinatorial optimization and local weak convergence. *Probability on discrete structures*, Volume 110, pp. 1-72.
- [2] Barbour, A., D. and Utev, S., (2004), Approximating the Reed-Frost epidemic process. *Stochastic Processes and their Applications*, Volume 113, pp. 173-197.
- [3] Draief, M., Ganesh, A. and Massoulié, L., (2008), Thresholds for virus spread on networks. *The Annals of Applied Probability*, Volume 18, No. 2, pp. 359-378.
- [4] Janson, S., Luczak, T. and Ruciński, N., (2000), *Random graphs*, John Wiley and Sons., New York.

- [5] Kermack, W. and MacKendrick, A., (1927), A contribution to mathematical theory of epidemics. *Proceedings of the Royal Society of London, Series A, Containing Papers of a Mathematical and Physical Character*, Volume 115, Issue 772, pp. 700-721
- [6] Lefevre, C. and Utev, S. (1995), Poisson approximation for the final state of a generalized epidemic process, *The Annals of Probability*, Volume 23, pp. 1139-1162.
- [7] Weaver N., Paxson, V., Staniford, S. and Cunningham, R. (2003), A taxonomy of computer worms, *In ACM Workshop on Rapid Malcode (WORM)* 11-18. ACM Press, New York.
- [8] Wästlund J. (2011), Replica symmetry of the minimum matching, *to appear in Annals of Mathematics*.
- [9] Athreya, K. B. and Ney, P. E. (2004), *Branching processes*, Dover Publications Inc., New York.
- [10] Grimmett, G. (1999), *Percolation (Second Edition)*, Springer-Verlag, Berlin.
- [11] Bollobás, B. (2001), *Random Graphs (Second Edition)*, Cambridge University Press, Cambridge.