ON THE EXPECTED TOTAL NUMBER OF INFECTIONS FOR VIRUS SPREAD ON A FINITE NETWORK

BY ANTAR BANDYOPADHYAY AND FARKHONDEH SAJADI

Indian Statistical Institute, Delhi Centre

In this work we consider a simple SIR infection spread model on a finite population of n agents represented by a finite graph G. Starting with a fixed set of initial infected vertices the infection spreads in discrete time steps, where each infected vertex tries to infect its neighbors with a fixed probability $\beta \in (0, 1)$, independently of others. It is assumed that each infected vertex dies out after an unit time and the process continues till all infected vertices die out. This model was first studied by [*Ann. Appl. Probab.* **18** (2008) 359–378]. In this work we find a simple lower bound on the expected number of ever infected vertices using *breath-first search* algorithm and show that it asymptotically performs better for a fairly large class of graphs than the upper bounds obtained in [*Ann. Appl. Probab.* **18** (2008) 359–378]. As a product we also derive the asymptotic value of the expected number of the ever infected vertices when the underlying graph is the random r-regular graph and $\beta < \frac{1}{r-1}$.

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 *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.

In this work we consider a simple virus spread model on a finite network of agents, and our goal is to determine an approximation of the total amount of infection without specifying the underlying network. The model we consider is a particular *susceptible infected removed* (*SIR*) model, which was first introduced in this context by Draief, Ganesh and Massoulié [7]. 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.

Like in [7], our goal is to study the total number of vertices that eventually become infected (and hence removed) without specifying the underlying network. In [7], the authors derived an explicit upper bound of the expected number of vertices ever infected which depends on both the size of the network as well as the infection rate $\beta \in (0, 1)$. This bound also needed an assumption of "small" value for β. Unfortunately, the work [7] did not provide any indication whether the derived upper bound is a good approximation of the quantity of interest. In this work we derive a simple lower bound of the expected number of vertices ever infected which works for every infection rate $0 < \beta < 1$. From application point of view, our lower bound is based on the *breadth-first search* (*BFS*) algorithm and hence is 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 it "*locally looks like a tree*," then our lower bound is asymptotically exact for "small" β, thus providing a good approximation when the network is "large." For such graphs G, the range we cover for β always includes the range in which the upper bound obtained in [7] holds and in all these cases, the upper bound over estimates the expected total number of infections. So our lower bound provides a better approximation.

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 vertices is assumed to be nonempty and all others are susceptible. The evolution of the epidemic is described by the following discrete time model:

- After a 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 [7], this is a simple model, falling in the class of models known as Reed–Frost models, where infection period is deterministic and is same for every vertex.

It is interesting to note here that, the model is essentially same as the i.i.d. Bernoulli bond percolation model [8] with parameter β. This is because the set

of ever infected (or removed) vertices is same as the union of connected open components of i.i.d. Bernoulli 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 and is known as the *binomial random graph*, also known as Erdős–Rényi random graph [4, 9]. Percolation on general finite graphs has also been studied in literature, particularly, the behavior of the giant component in the *critical* and *super-critical* regimes has been studied extensively [2, 5, 6, 10]. Our work falls in to the same domain, but we only study the size of the connected component of a given set of vertices and mainly for the *sub-critical* regime. More precisely, we derive asymptotic limit of the expected size of connected cluster of a set of vertices in the sub-critical regime for graphs with bounded degree and large girth. The *girth* of a finite graph is the size of the smallest cycle and thus the graphs we consider locally look like trees.

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.

We begin with a simple upper bound for $\mathbf{E}[Y^{G,I}]$ for graphs with bounded degree. This bound is used in the proofs of the main results stated later.

PROPOSITION 2.1.   *Suppose $G := (V, E)$ is a finite graph such that the degree of each vertex is bounded by $\Delta \geq 2$. Let $I \subseteq V$ be the set of vertices which are initially infected. Then for $\beta < \frac{1}{\Delta - 1}$ we have*

$$(2.1) \qquad \mathbf{E}[Y^{G,I}] \leq \frac{1 + \beta}{1 - (\Delta - 1)\beta}|I|.$$

We would like to point out that this upper bound is a better bound than that obtained in the Theorem 2.3 of [7]. Also it works for a larger set of parameter values.

PROOF OF PROPOSITION 2.1.   From definition of the model, we get

$$(2.2) \qquad \mathbf{E}[Y^{G,I}] \leq \sum_{v_0 \in I}\left(1 + \sum_{d=1}^{\infty} \pi_d(v_0)\beta^d\right),$$

where $\pi_d(v_0)$ is number of self-avoiding paths of length $d \geq 1$ starting at the vertex v_0. Since the maximum degree of G is bounded by Δ, so we must have $\pi_d(v_0) \leq \Delta(\Delta - 1)^{d-1}$ for all $d \geq 1$ and v_0. This completes the proof.   □

In the following subsection 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 Section 2.2.

In both cases, our results relay on a specific search algorithm, known as *breadth-first search* (*BFS*). For sake of completeness we briefly describe the algorithm here.

> Given a finite graph with a linear ordering of its vertices and a starting vertex called the root, the BFS algorithm begins at the root and inspects all the neighboring vertices. Then for each of those neighboring vertices in turn, it inspects their neighboring vertices which were unvisited, and continues this way till there are no unvisited vertices.

It is worth noting that the BFS algorithm produces a spanning tree of the given graph, but is not necessarily unique. The BFS spanning tree depends on the choice of the root, and it may also depend on the ordering of the vertices. Note that if the graph is a tree, then the BFS spanning tree is unique, and it is just the graph itself.

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

THEOREM 2.1. *Let $G := (V, E)$ 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*

$$(2.3) \qquad \mathbf{E}\big[Y^{T,\{v_0\}}\big] \leq \mathbf{E}\big[Y^{G,\{v_0\}}\big] \qquad \text{for all } 0 < \beta < 1.$$

Moreover, if $\mathcal{T}$ is a BFS spanning tree of the connected component of v_0 rooted at v_0, then

$$(2.4) \qquad \mathbf{E}\big[Y^{T,\{v_0\}}\big] \leq \mathbf{E}\big[Y^{\mathcal{T},\{v_0\}}\big] \leq \mathbf{E}\big[Y^{G,\{v_0\}}\big] \qquad \text{for all } 0 < \beta < 1.$$

PROOF.   Suppose T is a spanning subtree of G with root v_0. Then by the standard coupling we get

$$\mathbf{E}\big[Y^{T,\{v_0\}}\big] \leq \mathbf{E}\big[Y^{G,\{v_0\}}\big],$$

for any $0 < \beta < 1$.

To prove 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 $v \in V$. Thus

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

as $0 < \beta < 1$.   □

Let $\mathrm{LB}^{G,\{v_0\}} := \mathbf{E}[Y^{\mathcal{T},\{v_0\}}]$ 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 2.1 we get that

$$(2.5) \qquad \mathrm{LB}^{G,\{v_0\}} = \sum_{v \in V} \beta^{d_G(v,v_0)},$$

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 vertex.

For our next result we use the following graph theoretic notations. Given a graph G and a fixed vertex v_0 of G and a positive integer $d \geq 1$, we denote by $N_d(G, v_0)$, the induced sub-graph of G with vertices which are at a *graph distance* at most d from v_0.

THEOREM 2.2. *Let G_n be a connected graph on n vertices and $\{(G_n, v_0^n)\}_{n \geq 1}$ be a sequence of rooted graphs with roots $\{v_0^n\}_{n \geq 1}$ such that there exists a sequence $\alpha_n = \Omega(\log n)$ 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 $0 < \beta < \beta_0$,*

$$(2.6) \qquad \left| \mathbf{E}[Y^{G_n,\{v_0^n\}}] - \mathrm{LB}^{G_n,\{v_0^n\}} \right| \longrightarrow 0 \qquad as\ n \to \infty$$

and therefore $\dfrac{\mathbf{E}[Y^{G_n,\{v_0^n\}}]}{\mathrm{LB}^{G_n,\{v_0^n\}}} \longrightarrow 1$ *as* $n \to \infty$.

PROOF. Let $\mathcal{T}_n$ be a BFS spanning tree rooted at v_0^n of the graph G_n and as defined earlier let $\mathrm{LB}^{G_n,\{v_0^n\}} = \mathbf{E}[Y^{\mathcal{T}_n,\{v_0^n\}}]$. Denote $\partial_{\alpha_n}^* N_{\alpha_n}(G_n, v_0^n)$ the set of infected vertices in G_n after α_n units of time starting with one infected vertex v_0^n. Then

$$
\begin{aligned}
\mathrm{LB}^{G_n,\{v_0^n\}} &\leq \mathbf{E}[Y^{G_n,\{v_0^n\}}] \\
&\leq \mathbf{E}[Y^{N_{\alpha_n}(G_n,v_0^n),\{v_0^n\}}] + n\mathbf{E}[|\partial_{\alpha_n}^* N_{\alpha_n}(G_n, v_0^n)|] \\
&\leq \mathbf{E}[Y^{N_{\alpha_n}(G_n,v_0^n),\{v_0^n\}}] + n^2 \beta^{\alpha_n} \\
&\leq \mathrm{LB}^{G_n,\{v_0^n\}} + n^2 \beta^{\alpha_n}.
\end{aligned}
$$

(2.7)

The last inequality follows from the fact that by assumption $N_{\alpha_n}(G_n, v_0^n)$ is a tree and hence is a subtree of $\mathcal{T}_n$. This proves (2.6) since by assumption $\alpha_n = \Omega(\log n)$. The last part of the theorem follows from the fact that $\mathrm{LB}^{G_n,\{v_0^n\}} \geq 1$. $\square$

Although the assumption in the above theorem may seem to be very restrictive, it is satisfied in many examples. 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.

Before we state our next result we recall that 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*) [1], and we write $(G_n, v_0^n) \overset{\text{l.w.c.}}{\longrightarrow} (G_\infty, v_0^\infty)$, if for any $d \geq 1$,

$$(2.8) \qquad \mathbf{P}(N_d(G_n, v_0^n) \cong N_d(G_\infty, v_0^\infty)) \longrightarrow 1 \qquad \text{as } n \to \infty.$$

THEOREM 2.3.   *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 a fixed constant, namely $\Delta \geq 2$. Suppose there is a rooted deterministic or random tree $\mathscr{T}$ with root ρ such that*

$$(2.9) \qquad\qquad (G_n, v_0^n) \overset{\text{l.w.c.}}{\longrightarrow} (\mathscr{T}, \rho) \qquad \text{as } n \to \infty.$$

Let $\mathrm{LB}^{G_n, \{v_0^n\}} := \mathbf{E}[Y^{\mathcal{T}_n, \{v_0^n\}}]$ *where* $\mathcal{T}_n$ *is a BFS spanning tree rooted at* v_0^n *of the graph* G_n*. Then for $\beta < \frac{1}{\Delta - 1}$*

$$(2.10) \qquad\qquad (\mathbf{E}[Y^{G_n, \{v_0^n\}}] - \mathrm{LB}^{G_n, \{v_0^n\}}) \longrightarrow 0 \qquad \text{as } n \to \infty.$$

Moreover for $\beta < \frac{1}{\Delta - 1}$ we have

$$(2.11) \qquad \lim_{n \to \infty} \mathrm{LB}^{G_n, \{v_0^n\}} = \lim_{n \to \infty} \mathbf{E}[Y^{G_n, \{v_0^n\}}] = \mathbf{E}[Y^{\mathscr{T}, \{\rho\}}] < \infty.$$

PROOF.   Let $\mathcal{T}_n$ be a BFS spanning tree rooted at v_0^n of the graph G_n, and also as defined earlier let $\mathrm{LB}^{G_n, \{v_0^n\}} = \mathbf{E}[Y^{\mathcal{T}_n, \{v_0^n\}}]$. Fix $d \geq 1$ and E_n be the event $[N_d(G_n, v_0^n) \cong N_d(\mathscr{T}, \rho)]$. Note that E_n depends on d. Now from Theorem 2.1,

$$(2.12) \qquad \mathrm{LB}^{G_n, \{v_0^n\}} \leq \mathbf{E}[Y^{G_n, \{v_0^n\}}] = \mathbf{E}[Y^{G_n, \{v_0^n\}} \mathbf{1}_{E_n}] + \mathbf{E}[Y^{G_n, \{v_0^n\}} \mathbf{1}_{E_n^c}].$$

Now under our assumption, the degree of any vertex of G_n is bounded by Δ and $\beta < \frac{1}{\Delta - 1}$, so using inequality (2.1), we get

$$(2.13) \qquad\qquad \mathbf{E}[Y^{G_n, \{v_0^n\}} \mathbf{1}_{E_n^c}] \leq \frac{1 + \beta}{1 - (\Delta - 1)\beta} \mathbf{P}(E_n^c).$$

Further note that if E_n occurs, $N_d(G_n, v_0^n)$ is a tree rooted at v_0^n and hence it is then a sub-tree of $\mathcal{T}_n$. Thus if E_n occurs, $N_d(G_n, v_0^n) = N_d(\mathcal{T}_n, v_0^n)$. Like earlier, let $\partial_d^* N_d(\mathcal{T}_n, v_0^n)$ be the set of infected vertices in $\mathcal{T}_n$ after d units of time starting with one infected vertex v_0^n. Then we have

$$\mathbf{E}[Y^{G_n, \{v_0^n\}} \mathbf{1}_{E_n}] \leq \mathbf{E}[Y^{N_d(\mathcal{T}_n, v_0^n), \{v_0^n\}} \mathbf{1}_{E_n}] + \mathbf{E}[Y^{G_n, \partial_d^* N_d(\mathcal{T}_n, v_0^n)} \mathbf{1}_{E_n}]$$

$$(2.14) \qquad\qquad \leq \mathrm{LB}^{G_n, \{v_0^n\}} + \frac{1 + \beta}{1 - (\Delta - 1)\beta} \mathbf{E}[|\partial_d^* N_d(\mathcal{T}_n, v_0^n)|]$$

$$\leq \mathrm{LB}^{G_n, \{v_0^n\}} + \frac{\beta(1 + \beta)\Delta(\beta(\Delta - 1))^{d-1}}{1 - (\Delta - 1)\beta}.$$

For the second inequality, we use (2.1). For the last inequality, note that there are at most $\Delta(\Delta - 1)^{d-1}$ self-avoiding paths of length d from v_0^n and each path has probability β^d of infections occurring all along the path. Therefore $\mathbf{E}[|\partial_d^* N_d(\mathcal{T}_n, v_0^n)|] \leq \beta \Delta(\beta(\Delta - 1))^{d-1}$.

Now under assumption (2.9), we have $\lim_{n \to \infty} \mathbf{P}(E_n^c) = 0$, so we conclude that for any $d \geq 1$,

$$0 \leq \limsup_{n \to \infty}(\mathbf{E}[Y^{G_n, \{v_0^n\}}] - \mathrm{LB}^{G_n, \{v_0^n\}}) \leq \frac{\beta(1 + \beta)\Delta(\beta(\Delta - 1))^{d-1}}{1 - (\Delta - 1)\beta}.$$

This proves (2.10) by taking $d \to \infty$ as $\beta < \frac{1}{\Delta-1}$.

Now for proving (2.11), we first observe that from (2.9) the degree of any vertex of $\mathcal{T}$ is also bounded by Δ. So using (2.1), we get that for $\beta < \frac{1}{\Delta-1}$

$$\mathbf{E}[Y^{N_d(\mathcal{T}, \rho), \{\rho\}}] \leq \frac{1 + \beta}{1 - (\Delta - 1)\beta}.$$

Moreover from the definition, $Y^{N_d(\mathcal{T}, \rho), \{\rho\}} \uparrow Y^{\mathcal{T}, \{\rho\}}$ as $d \to \infty$. So by the monotone convergence theorem, we have

$$(2.15) \qquad \lim_{d \to \infty} \mathbf{E}[Y^{N_d(\mathcal{T}, \rho), \{\rho\}}] = \mathbf{E}[Y^{\mathcal{T}, \{\rho\}}] \leq \frac{1 + \beta}{1 - (\Delta - 1)\beta} < \infty.$$

Fix $\varepsilon > 0$, since $\beta < \frac{1}{\Delta-1}$, so we can find $d \geq 1$ such that

$$(2.16) \qquad \left|\mathbf{E}[Y^{\mathcal{T}, \{\rho\}}] - \mathbf{E}[Y^{N_d(\mathcal{T}, \rho), \{\rho\}}]\right| < \varepsilon$$

and

$$(2.17) \qquad \frac{\beta(1 + \beta)\Delta(\beta(\Delta - 1))^{d-1}}{1 - (\Delta - 1)\beta} < \varepsilon.$$

Further, as the degree of any vertex of $\mathcal{T}$ is bounded by Δ, so arguing similarly to the derivation of the equation (2.13), we conclude

$$(2.18) \qquad \mathbf{E}[Y^{N_d(\mathcal{T}, \rho), \{\rho\}} \mathbf{1}_{E_n^c}] \leq \frac{1 + \beta}{1 - (\Delta - 1)\beta} \mathbf{P}(E_n^c).$$

Further combining (2.13) and the first inequality of (2.14), we get

$$
\begin{aligned}
&\left|\mathbf{E}[Y^{G_n, \{v_0^n\}}] - \mathbf{E}[Y^{N_d(G_n, v_0^n), \{v_0^n\}} \mathbf{1}_{E_n}]\right| \\
(2.19) \qquad &\leq \frac{\beta(1 + \beta)\Delta(\beta(\Delta - 1))^{d-1}}{1 - (\Delta - 1)\beta} + \frac{1 + \beta}{1 - (\Delta - 1)\beta} \mathbf{P}(E_n^c) \\
&\leq \varepsilon + \frac{1 + \beta}{1 - (\Delta - 1)\beta} \mathbf{P}(E_n^c),
\end{aligned}
$$

where the last equality follows from (2.17). Finally,

$$
\begin{aligned}
\left|\mathbf{E}\big[Y^{G_n,\{v_0^n\}}\big] - \mathbf{E}\big[Y^{\mathscr{T},\{\rho\}}\big]\right| &\leq \left|\mathbf{E}\big[Y^{G_n,\{v_0^n\}}\big] - \mathbf{E}\big[Y^{N_d(G_n,v_0^n),\{v_0^n\}}\mathbf{1}_{E_n}\big]\right| \\
&\quad + \left|\mathbf{E}\big[Y^{N_d(G_n,v_0^n),\{v_0^n\}}\mathbf{1}_{E_n}\big] - \mathbf{E}\big[Y^{N_d(\mathscr{T},\rho),\{\rho\}}\big]\right| \\
&\quad + \left|\mathbf{E}\big[Y^{N_d(\mathscr{T},\rho),\{\rho\}}\big] - \mathbf{E}\big[Y^{\mathscr{T},\{\rho\}}\big]\right| \\
&\leq 2\varepsilon + \frac{2(1+\beta)}{1-(\Delta-1)\beta}\mathbf{P}(E_n^c),
\end{aligned}
$$

where the last inequality follows from equations (2.16), (2.17), (2.18) and (2.19) and also observing the fact that

$$
\mathbf{E}\big[Y^{N_d(G_n,v_0^n),\{v_0^n\}}\mathbf{1}_{E_n}\big] = \mathbf{E}\big[Y^{N_d(\mathscr{T},\rho),\{\rho\}}\mathbf{1}_{E_n}\big].
$$

Now under our assumption (2.9) we have $\mathbf{P}(E_n) \longrightarrow 1$. So we conclude that

$$
(2.20) \qquad \lim_{n\to\infty} \mathbf{E}\big[Y^{G_n,\{v_0^n\}}\big] = \mathbf{E}\big[Y^{\mathscr{T},\{\rho\}}\big].
$$

This completes the proof using (2.10). $\square$

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

THEOREM 2.4. *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 $\beta < \frac{1}{r-1}$,*

$$
(2.21) \qquad \lim_{n\to\infty} \mathbf{E}\big[Y^{G_n,\{v_0^n\}}\big] = \frac{1+\beta}{1-(r-1)\beta}.
$$

We note that in this case, the upper bound given in [7] is $\frac{1}{1-r\beta}$ when $\beta < \frac{1}{r}$ which is strictly bigger than the exact answer given in (2.21).

PROOF OF THEOREM 2.4. It is known [1, 9] 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

$$
(2.22) \qquad \big(G_n, v_0^n\big) \overset{\text{l.w.c.}}{\longrightarrow} (\mathbb{T}_r, \phi),
$$

where $\mathbb{T}_r$ is the infinite r-regular tree with root say ϕ. Now $\mathbf{E}[Y^{\mathbb{T}_r,\phi}] = 1 + r\beta\mathbf{E}[S_{r,\beta}]$ where $S_{r,\beta}$ is the total size of a sub-critical Galton–Watson branching process with progeny distribution Binomial$(r-1, \beta)$. The result then follows from Theorem 2.3 and the fact $\mathbf{E}[S_{r,\beta}] = \frac{1}{1-(r-1)\beta}$ [3]. $\square$

2.2. *Starting with more than one infected vertex.* Now suppose instead of one infected vertex, we start with k infected vertices given by $I := \{v_{0,1}, v_{0,2}, \ldots, v_{0,k}\}$. The following theorem gives a lower bound similar to that of Theorem 2.1.

THEOREM 2.5. *Let G be an arbitrary finite graph and $I := \{v_{0,j}\}_{j=1}^{k}$ be a fixed set of k vertices. Let 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*

$$(2.23) \qquad \mathbf{E}[Y^{T,I}] \leq \mathbf{E}[Y^{G,I}] \qquad \text{for all } 0 < \beta < 1.$$

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*

$$(2.24) \qquad \mathbf{E}[Y^{T,I}] \leq \mathbf{E}[Y^{\mathcal{T},I}] \leq \mathbf{E}[Y^{G,I}] \qquad \text{for all } 0 < \beta < 1.$$

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 $\{v_{0,1}, v_{0,2}, \ldots, v_{0,k}\}$, that are obtained through the *breath-first-search* algorithm, starting at some vertex $v \in I$ and assuming that all the vertices $\{v_{0,1}, v_{0,2}, \ldots, v_{0,k}\}$ are at the same level. Alternately, 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 $v_{0,1}, v_{0,2}, \ldots, v_{0,k}$ through k "artificial" edges, and we perform the BFS algorithm on G^* starting with the vertex v^*, to obtain 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 $\{v_{0,1}, v_{0,2}, \ldots, v_{0,k}\}$ is given by the forest $\mathcal{T}^* \setminus \{v^*\}$. This alternate description is quite useful in practice. Note that if $\{\mathcal{T}_i\}_{1 \leq i \leq k}$ are the k connected components, rooted, respectively, at $\{v_{0,1}, v_{0,2}, \ldots, v_{0,k}\}$ of $\mathcal{T}$, a breath-first-search spanning forest of the connected components of G containing the vertices in I, then the following identity holds for every $\beta \in (0, 1)$:

$$(2.25) \qquad \mathbf{E}[Y^{\mathcal{T},I}] = \sum_{i=1}^{k} \mathbf{E}[Y^{\mathcal{T}_i,I}] = \frac{\mathbf{E}[Y^{\mathcal{T}^*,\{v^*\}}] - 1}{\beta}.$$

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

We write $\mathrm{LB}^{G,I}$ for $\mathbf{E}[Y^{\mathcal{T},I}]$ which is the lower bound of $\mathbf{E}[Y^{G,I}]$ for starting with k infected vertices given by I. Observe that from equation (2.25) we can write

$$(2.26) \qquad \mathrm{LB}^{G,I} = \sum_{i=1}^{k} \mathbf{E}[Y^{\mathcal{T}_i,I}],$$

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 (2.26) uses a specific choice of $\mathcal{T}$.

THEOREM 2.6. *Let* $\{(G_n, I_n)\}_{n \geq 1}$ *be a sequence of graphs where each* G_n *has k-roots given by the set* $I_n := \{v_{0,1}^n, v_{0,2}^n, \ldots, v_{0,k}^n\}$ *such that there exists a sequence* $\alpha_n = \Omega(\log n)$ *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* $0 < \beta < \beta_0$,

$$(2.27) \qquad \left| \mathbf{E}\big[Y^{G_n, I_n}\big] - \text{LB}^{G_n, I_n} \right| \longrightarrow 0 \qquad as\ n \to \infty$$

and therefore $\frac{\mathbf{E}[Y^{G_n, I_n}]}{\text{LB}^{G_n, I_n}} \longrightarrow 1$ *as* $n \to \infty$.

The proof of this result is similar to that of Theorem 2.2 and follows from the identity (2.25). The details are thus omitted.

Our next result is parallel to the Theorem 2.3 which needs a generalization of the concept of local weak convergence which was introduced by Wästlund [11].

We will say a sequence of random or deterministic graphs $\{G_n\}_{n \geq 1}$ with k-roots given by $I_n := \{v_{0,1}^n, v_{0,2}^n, \ldots, v_{0,k}^n\}$, $n \geq 1$ converges to a random or deterministic graph G_∞ with k-roots say $I_\infty := \{v_{0,1}^\infty, v_{0,2}^\infty, \ldots, v_{0,k}^\infty\}$ 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 $d \geq 1$,

$$(2.28) \qquad \mathbf{P}\big(N_d\big(G_n, v_{0,j}^n\big) \cong N_d\big(G_\infty, v_{0,j}^\infty\big) \text{ for all } 1 \leq j \leq k\big) \longrightarrow 1$$
$$as\ n \to \infty.$$

Note that for a sequence deterministic graphs, (2.28) means that the event occurs for "large" enough n.

THEOREM 2.7. *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* $I_n := \{v_{0,1}^n, v_{0,2}^n, \ldots, v_{0,k}^n\}$ *and maximum degree of each* G_n *is bounded by a fixed constant, namely* $\Delta \geq 2$. *Suppose* $\mathcal{T} := \bigcup_{j=1}^{k} \mathcal{T}_j$ *is a forest with k-rooted tress with roots* $I_\infty := \{\phi_1, \phi_2, \ldots, \phi_k\}$. *We assume that*

$$(2.29) \qquad (G_n, I_n) \xrightarrow{\text{l.w.c.}} (\mathcal{T}, I_\infty) \qquad as\ n \to \infty.$$

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

$$(2.30) \qquad \big(\mathbf{E}\big[Y^{G_n, I_n}\big] - \text{LB}^{G_n, I_n}\big) \longrightarrow 0,$$

as $n \to \infty$. *Moreover for* $\beta < \frac{1}{\Delta - 1}$ *we have*

$$(2.31) \qquad \lim_{n \to \infty} \text{LB}^{G_n, I_n} = \lim_{n \to \infty} \mathbf{E}\big[Y^{G_n, I_n}\big] = \mathbf{E}\big[Y^{\mathcal{T}, I_\infty}\big] = \sum_{j=1}^{k} \mathbf{E}\big[Y^{\mathcal{T}_j, \{\phi_j\}}\big].$$

PROOF. For each $n \geq 1$, as 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 $\{\phi_1, \phi_2, \ldots, \phi_k\}$. Then our assumption of local weak convergence (2.29) is equivalent to

$$(2.32) \qquad (G_n^*, v_n^*) \xrightarrow{\text{l.w.c.}} (\mathcal{T}^*, \phi^*).$$

This together with relation (2.25) and Theorem 2.3 completes the proof. □

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 (2.31) can be reformulated as

$$(2.33) \qquad \lim_{n \to \infty} \text{LB}^{G_n, I_n} = \lim_{n \to \infty} \mathbf{E}[Y^{G_n, I_n}] = \mathbf{E}[Y^{\mathcal{T}, I_\infty}] = k\mathbf{E}[Y^{\mathcal{T}_1, \{\phi_1\}}].$$

As in the case of starting with one infected vertex, the following theorem is an immediate application of the above results.

THEOREM 2.8. *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 $I_n := \{v_{0,1}^n, v_{0,2}^n, \ldots, v_{0,k}^n\}$ be k uniformly and independently selected vertices of G_n. Then for $\beta < \frac{1}{r-1}$,*

$$(2.34) \qquad \lim_{n \to \infty} \mathbf{E}[Y^{G_n, I_n}] = k\frac{1 + \beta}{1 - (r-1)\beta}.$$

PROOF. Since the vertices in I_n are selected uniformly at random, from [1] we have

$$(2.35) \qquad (G_n, I_n) \xrightarrow{\text{l.w.c.}} (\mathcal{T}_r, I_\infty),$$

where $I_\infty := \{\phi_1, \phi_2, \ldots, \phi_k\}$ and $\mathcal{T}_r$ is a forest with k infinite r-regular tree with roots in I_∞. The result then follows from Theorems 2.7 and 2.4. □

Once again we note that in this case, the upper bound $\frac{k}{1 - r\beta}$ given in [7] for $\beta < \frac{1}{r}$, is strictly bigger than the exact answer given in (2.34), and the gap increases with k, the initial number of infections.

3. 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 of this expectation. Although an approximation from above with an upper bound would be a more conservative and hence useful from practical point view. Unfortunately, the known upper bounds (e.g., the ones derived in [7]) hold only for "small" values of the parameter β and often are much larger than the exact quantity. For an arbitrary finite network, we have obtained a lower bound of the

expectation of the number of vertices ever infected for any value of the parameter β which is computable through the breadth-first search algorithm. Theorems 2.2, 2.3, 2.6 and 2.7 show that this lower bound is asymptotically exact for a large class of graphs when β value is "small," which always includes the values of β for which the upper bounds in [7] are defined.

As discussed in the Introduction in disguise, the work also provides lower bound for the expected size of the cluster of i.i.d. Bernoulli bond percolation model on finite graphs. Theorems 2.2, 2.3, 2.6 and 2.7 show that this bound is asymptotically exact, but only in the *sub-critical* regime, for graphs with bounded degree with large girth.

REFERENCES

[1] ALDOUS, D. and STEELE, J. M. (2004). The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on Discrete Structures*. *Encyclopaedia Math. Sci.* **110** 1–72. Springer, Berlin. MR2023650

[2] ALON, N., BENJAMINI, I. and STACEY, A. (2004). Percolation on finite graphs and isoperimetric inequalities. *Ann. Probab.* **32** 1727–1745. MR2073175

[3] ATHREYA, K. B. and NEY, P. E. (2004). *Branching Processes*. Dover Publications, Inc., Mineola, NY. MR2047480

[4] BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. Cambridge Univ. Press, Cambridge. MR1864966

[5] BORGS, C., CHAYES, J. T., VAN DER HOFSTAD, R., SLADE, G. and SPENCER, J. (2005). Random subgraphs of finite graphs. I. The scaling window under the triangle condition. *Random Structures Algorithms* **27** 137–184. MR2155704

[6] BORGS, C., CHAYES, J. T., VAN DER HOFSTAD, R., SLADE, G. and SPENCER, J. (2005). Random subgraphs of finite graphs. II. The lace expansion and the triangle condition. *Ann. Probab.* **33** 1886–1944. MR2165583

[7] DRAIEF, M., GANESH, A. and MASSOULIÉ, L. (2008). Thresholds for virus spread on networks. *Ann. Appl. Probab.* **18** 359–378. MR2398760

[8] GRIMMETT, G. (1999). *Percolation*, 2nd ed. Springer, Berlin. MR1707339

[9] JANSON, S., ŁUCZAK, T. and RUCINSKI, A. (2000). *Random Graphs*. Wiley, New York. MR1782847

[10] KOZMA, G. and NACHMIAS, A. (2011). A note about critical percolation on finite graphs. *J. Theoret. Probab.* **24** 1087–1096. MR2851246

[11] WÄSTLUND, J. (2012). Replica symmetry of the minimum matching. *Ann. of Math.* (2) **175** 1061–1091. MR2912702

THEORETICAL STATISTICS AND MATHEMATICS UNIT
INDIAN STATISTICAL INSTITUTE, DELHI CENTRE
7 S. J. S. SANSANWAL MARG
NEW DELHI 110016
INDIA
E-MAIL: antar@isid.ac.in
            farkhondeh.sajadi@gmail.com
URL: http://www.isid.ac.in/~antar