What I Tell You Three Times Is True: Bootstrap Percolation in Small Worlds*

Hamed Amini** and Nikolaos Fountoulakis***

EPFL and University of Birmingham
hamed.amini@epfl.ch, N.Fountoulakis@bham.ac.uk

Abstract. A bootstrap percolation process on a graph G is an "infection" process which evolves in rounds. Initially, there is a subset of infected nodes and in each subsequent round each uninfected node which has at least r infected neighbours becomes infected and remains so forever. The parameter $r \geqslant 2$ is fixed.

We analyse this process in the case where the underlying graph is an inhomogeneous random graph, which exhibits a power-law degree distribution, and initially there are $a(n)$ randomly infected nodes. The main focus of this paper is the number of vertices that will have been infected by the end of the process. The main result of this work is that if the degree sequence of the random graph follows a power law with exponent β, where $2 < \beta < 3$, then a sublinear number of initially infected vertices is enough to spread the infection over a linear fraction of the nodes of the random graph, with high probability.

More specifically, we determine explicitly a critical function $a_c(n)$ such that $a_c(n) = o(n)$ with the following property. Assuming that n is the number of vertices of the underlying random graph, if $a(n) \ll a_c(n)$, then the process does not evolve at all, with high probability as n grows, whereas if $a(n) \gg a_c(n)$, then there is a constant $\varepsilon > 0$ such that, with high probability, the final set of infected vertices has size at least εn. This behaviour is in sharp contrast with the case where the underlying graph is a $G(n,p)$ random graph with $p = d/n$. Recent results of Janson, Łuczak, Turova and Vallier have shown that if the number of initially infected vertices is sublinear, then with high probability the size of the final set of infected vertices is approximately equal to $a(n)$. That is, essentially there is lack of evolution of the process.

It turns out that when the maximum degree is $o(n^{1/(\beta-1)})$, then $a_c(n)$ depends also on r. But when the maximum degree is $\Theta(n^{1/(\beta-1)})$, then $a_c(n) = n^{\frac{\beta-2}{\beta-1}}$.

Keywords: bootstrap percolation, contagion, power-law random graphs.

* L. Carroll *The Hunting of the Snark.*

** Financial support from the Austrian Science Fund (FWF) project P21709 is gratefully acknowledged.

*** This research has been supported by a Marie Curie Intra-European Research Fellowship PIEF-GA-2009-255115 hosted by the Max-Planck-Institut für Informatik.

P.W. Goldberg and M. Guo (Eds.): WINE 2012, LNCS 7695, pp. 462–474, 2012.

1 Introduction

Models for the processes by which new ideas and new behaviors propagate through a population have been studied in a number of domains, including the epidemiology, political science, agriculture, finance and the effects of word of mouth (also known as viral marketing) in the promotion of new products. An idea or innovation appears (for example, the use of a new technology among college students) and it can either die out quickly or make significant advances into the population. The hypothesis of viral marketing is that by initially targeting a few influential members of the network (e.g., by giving them free samples of the product), we can trigger a cascade of influence by which friends will recommend the product to other friends, and many individuals will ultimately try it. But how should we choose the few key individuals to use for seeding this process? This problem is known as "the influence maximization problem"; hardness results have been obtained in [29], [30] and there is a large literature on this topic (see for example [31] and the references therein). However, in most practical cases, the structure of the underlying network is not known and then one has to initially target the popular and attractive individuals with many connections.

In this paper, we consider a simple model of diffusion, known as "bootstrap percolation model". Bootstrap percolation was introduced by Chalupa, Leath and Reich [13] in 1979 in the context of magnetic disordered systems and has been re-discovered since then by several authors mainly due to its connections with various physical models. A *bootstrap percolation process* with *activation threshold* an integer $r \geqslant 2$ on a graph $G = G(V, E)$ is a deterministic process which evolves in rounds. Every vertex has two states: it is either *infected* or *uninfected*. Initially, there is a subset $\mathcal{A}_0 \subseteq V$ which consists of infected vertices, whereas every other vertex is uninfected. This set can be selected either deterministically or randomly. Subsequently, in each round, if an uninfected vertex has at least r of its neighbours infected, then it also becomes infected and remains so forever. This is repeated until no more vertices become infected. We denote the final infected set by $\mathcal{A}_f$.

Bootstrap percolation processes (and extensions) have been used as models to describe several complex phenomena in diverse areas, from jamming transitions [27] and magnetic systems [24] to neuronal activity [3], [26] and spread of defaults in banking systems (see e.g. [4] with a more refined model). A short survey regarding applications of bootstrap percolation processes can be found in [1].

In the context of real-world networks and in particular in social networks, a bootstrap percolation process can be thought of as a primitive model for the spread of ideas or new trends within a set of individuals which form a network. Each of them has a threshold r and $\mathcal{A}_0$ corresponds to the set of individuals who initially are "infected" with a new belief. If for an "uninfected" individual at least r of its acquaintances have adopted the new belief, then this individual adopts it as well. Bootstrap percolation processes have also been studied on a variety of graphs, such as trees [8], [18], grids [12], [20], [7], [6], hypercubes [5], as well as on several distributions of random graphs [9], [22], [2].

More than a decade ago, Faloutsos et al. [17] observed that the Internet exhibits a *power-law* degree distribution, meaning that the proportion of vertices of degree k scales like $k^{-\beta}$, for all sufficiently large k, and some $\beta > 2$. In particular, the work of Faloutsos et al. [17] suggested that the degree distribution of the Internet at the router level follows a power law with $\beta \approx 2.6$. Kumar et al. [23] also provided evidence on the degree distribution of the World Wide Web viewed as a directed graph on the set of web pages, where a web page "points" to another web page if the former contains a link to the latter. They found that the indegree distribution follows a power law with exponent approximately 2.1, whereas the outdegree distribution follows also a power law with exponent close to 2.7. Other empirical evidence on real-world networks has provided examples of power law degree distributions with exponents between 2 and 3.

Thus, in the present work, we focus on the case where $2 < \beta < 3$. More specifically, the underlying random graph distribution we consider was introduced by Chung and Lu [14], who invented it as a general purpose model for generating graphs with a power-law degree sequence. Consider the vertex set $[n] := \{1, \ldots, n\}$. Every vertex $i \in [n]$ is assigned a positive weight w_i, and the pair $\{i, j\}$, for $i \neq j \in [n]$, is included in the graph as an edge with probability proportional to $w_i w_j$, independently of every other pair. Note that the expected degree of i is close to w_i. With high probability the degree sequence of the resulting graph follows a power law, provided that the sequence of weights follows a power law (see [28] for a detailed discussion). Such random graphs are also characterized as *ultra-small worlds*, due to the fact that the typical distance of two vertices that belong to the same component is $O(\log \log n)$ – see [15] or [28].

Regarding the initial conditions of the bootstrap percolation process, our general assumption will be that the initial set of infected vertices $\mathcal{A}_0$ is chosen randomly among all subsets of vertices of a certain size.

The aim of this paper is to analyse the evolution of the bootstrap percolation process on such random graphs and, in particular, the typical value of the ratio $|\mathcal{A}_f|/|\mathcal{A}_0|$. The main finding of the present work is the existence of a critical function $a_c(n)$, which is sublinear, such that when $|\mathcal{A}_0|$ "crosses" $a_c(n)$ we have a sharp change on the evolution of the bootstrap percolation process. When $|\mathcal{A}_0| \ll a_c(n)$, then typically the process does not evolve, but when $|\mathcal{A}_0| \gg a_c(n)$, then a linear fraction of vertices is eventually infected. Of course the non-trivial case here is when $|\mathcal{A}_0|$ is sublinear. What turns out to be the key to such a dissemination of the infection is the vertices of high weight. These are typically the vertices that have high degree in the random graph and, moreover, they form a fairly dense graph. We exploit this fact and show how this causes the spread of the infection to a linear fraction of the vertices (see Theorem 2 below). Interpreting this from the point of view of a social network, these vertices correspond to popular and attractive individuals with many connections – these are the *hubs* of the network. Our analysis sheds light to the role of these individuals in the infection process.

These results are in sharp contrast with the behaviour of the bootstrap percolation process in $G(n, p)$ random graphs, where every edge on a set of n vertices is included independently with probability p. Recently, Janson, Łuczak, Turova and Vallier [22] came up with a complete analysis of the bootstrap percolation process for various ranges of the probability p. Since the random graphs we consider have constant average degree, we focus on their findings regarding the range where $p = d/n$ and $d > 0$ is fixed. Among the findings of Janson et al. [22] (see Theorem 5.2 there) is that when $|\mathcal{A}_0| = o(n)$, then typically the process essentially does not evolve. More precisely, the ratio $|\mathcal{A}_f|/|\mathcal{A}_0|$ converges to 1 in probability – see below for the definition of this notion. In other words, the density of the initially infected vertices must be positive in order for the density of infected vertices to grow. We note that similar behavior to the case of $G(n,p)$ has been observed in the case of random regular graphs [9], and in random graphs with given vertex degrees constructed through the configuration model, studied by the first author in [2], when the sum of the square of degrees scales linearly with n, the size of the graph. The later case includes random graphs with power-law degree sequence with exponent $\beta > 3$. Our results imply that the two regimes $2 < \beta < 3$ and $\beta > 3$ have completely different behaviors.

Basic Notations. Let $\mathbb{R}^+$ be the set of positive real numbers. For non-negative sequences x_n and y_n, we describe their relative order of magnitude using Landau's $o(.)$ and $O(.)$ notation. We write $x_n = O(y_n)$ if there exist $N \in \mathbb{N}$ and $C > 0$ such that $x_n \leqslant C y_n$ for all $n \geqslant N$, and $x_n = o(y_n)$, if $x_n/y_n \to 0$, as $n \to \infty$. We also write $x_n \ll y_n$ when $x_n = o(y_n)$ and $x_n \gg y_n$ when $y_n = o(x_n)$.

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of real-valued random variables on a sequence of probability spaces $\{(\Omega_n, \mathbb{P}_n)\}_{n\in\mathbb{N}}$. If $c \in \mathbb{R}$ is a constant, we write $X_n \stackrel{p}{\to} c$ to denote that X_n *converges in probability to* c. That is, for any $\varepsilon > 0$, we have $\mathbb{P}_n(|X_n - c| > \varepsilon) \to 0$ as $n \to \infty$.
Let $\{a_n\}_{n\in\mathbb{N}}$ be a sequence of real numbers that tends to infinity as $n \to \infty$. We write $X_n = o_p(a_n)$, if $|X_n|/a_n$ *converges to 0 in probability.* Additionally, we write $X_n = O_p(a_n)$, to denote that for any positive-valued function $\omega(n) \to \infty$, as $n \to \infty$, we have $\mathbb{P}(|X_n|/a_n \geqslant \omega(n)) = o(1)$. If $\mathcal{E}_n$ is a measurable subset of Ω_n, for any $n \in \mathbb{N}$, we say that the sequence $\{\mathcal{E}_n\}_{n\in\mathbb{N}}$ occurs *asymptotically almost surely (a.a.s.)* if $\mathbb{P}(\mathcal{E}_n) = 1 - o(1)$, as $n \to \infty$.

Also, we denote by $\mathsf{Be}(p)$ a Bernoulli distributed random variable whose probability of being equal to 1 is p. The notation $\mathsf{Bin}(k,p)$ denotes a binomially distributed random variable corresponding to the number of successes of a sequence of k independent Bernoulli trials each having probability of success equal to p.

2 Models and Results

The random graph model that we consider is asymptotically equivalent to a model considered by Chung and Lu [15], and is a special case of the so-called *inhomogeneous random graph*, which was introduced by Söderberg [25] and was generalised and studied in great detail by Bollobás, Janson and Riordan in [11].

2.1 Inhomogeneous Random Graphs – The Chung-Lu Model

In order to define the model we consider for any $n \in \mathbb{N}$ the vertex set $[n] := \{1, \ldots, n\}$. Each vertex i is assigned a positive weight $w_i(n)$, and we will write $\mathbf{w} = \mathbf{w}(n) = (w_1(n), \ldots, w_n(n))$. We assume in the remainder that the weights are deterministic, and we will suppress the dependence on n, whenever this is obvious from the context. However, note that the weights could themselves be random variables; we will not treat this case here, although it is very likely that under suitable technical assumptions our results generalize to this case as well. For any $S \subseteq [n]$, set

$$W_S(\mathbf{w}) := \sum_{i \in S} w_i.$$

In our random graph model, the event of including the edge $\{i, j\}$ in the resulting graph is independent of the events of including all other edges, and equals

$$p_{ij}(\mathbf{w}) = \min\left\{\frac{w_i w_j}{W_{[n]}(\mathbf{w})}, 1\right\}. \quad (1)$$

This model was considered by Chung et al., for fairly general choices of $\mathbf{w}$, who studied in a series of papers [14–16] several typical properties of the resulting graphs, such as the average path length or the component distribution. We will refer to this model as the *Chung-Lu* model, and we shall write $CL(\mathbf{w})$ for a random graph in which each possible edge $\{i, j\}$ is included independently with probability as in (1). Moreover, we will suppress the dependence on $\mathbf{w}$, if it is clear from the context which sequence of weights we refer to.

Note that in a Chung-Lu random graph, the weights essentially control the *expected* degrees of the vertices. Indeed, if we ignore the minimization in (1), and also allow a loop at vertex i, then the expected degree of that vertex is $\sum_{j=1}^{n} w_i w_j / W_{[n]} = w_i$. In the general case, a similar asymptotic statement is true, unless the weights fluctuate too much. Consequently, the choice of $\mathbf{w}$ has a significant effect on the degree sequence of the resulting graph. For example, the authors of [15] choose $w_i = d\frac{\beta-2}{\beta-1}(\frac{n}{i+i_0})^{1/(\beta-1)}$, which typically results in a graph with a power-law degree sequence with exponent β, average degree d, and maximum degree proportional to $(n/i_0)^{1/(\beta-1)}$, where i_0 was chosen such that this expression is $O(n^{1/2})$. Our results will hold in a more general setting, where larger fluctuations around a "strict" power law are allowed, and also larger maximum degrees are possible, thus allowing a greater flexibility in the choice of the parameters.

2.2 Power-Law Degree Distributions

Following van der Hofstad [28], let us write for any $n \in \mathbb{N}$ and any sequence of weights $\mathbf{w} = (w_1(n), \ldots, w_n(n))$

$$F_n(x) = n^{-1} \sum_{i=1}^{n} \mathbf{1}[w_i(n) < x], \quad \forall x \in [0, \infty)$$

for the empirical distribution function of the weight of a vertex chosen uniformly at random. We will assume that F_n satisfies the following two conditions.

Definition 1. *We say that* $(F_n)_{n \geqslant 1}$ *is* regular*, if it has the following two properties.*

- **[Weak convergence of weight]** *There is a distribution function* $F : [0, \infty) \to [0, 1]$ *such that for all* x *at which* F *is continuous* $\lim_{n\to\infty} F_n(x) = F(x)$*;*
- **[Convergence of average weight]** *Let* W_n *be a random variable with distribution function* F_n*, and let* W_F *be a random variable with distribution function* F*. Then we have* $\lim_{n\to\infty} \mathbb{E}[W_n] = \mathbb{E}[W_F]$*.*

The regularity of $(F_n)_{n \geqslant 1}$ guarantees two important properties. Firstly, the weight of a random vertex is approximately distributed as a random variable that follows a certain distribution. Secondly, this variable has finite mean and therefore the resulting graph has bounded average degree. Apart from regularity, our focus will be on weight sequences that give rise to power-law degree distributions.

Definition 2. *We say that a regular sequence* $(F_n)_{n \geqslant 1}$ *is* of power law with exponent β*, if there are* $0 < \gamma_1 < \gamma_2$*,* $x_0 > 0$ *and* $0 < \zeta \leqslant 1/(\beta - 1)$ *such that for all* $x_0 \leqslant x \leqslant n^{\zeta}$

$$\gamma_1 x^{-\beta+1} \leqslant 1 - F_n(x) \leqslant \gamma_2 x^{-\beta+1},$$

and $F_n(x) = 0$ *for* $x < x_0$*, but* $F_n(x) = 1$ *for* $x > n^{\zeta}$*.*

Thus, we may assume that for $1 \leqslant i \leqslant n(1 - F_n(n^{\zeta}))$ we have $w_i = n^{\zeta}$, whereas for $(1 - F_n(n^{\zeta}))n < i \leqslant n$ we have $w_i = [1 - F_n]^{-1}(i/n)$, where $[1 - F_n]^{-1}$ is the generalized inverse of $1 - F_n$, that is, for $x \in [0, 1]$ we define $[1 - F_n]^{-1}(x) = \inf\{s \ : \ 1 - F_n(s) < x\}$. Note that according to the above definition, for $\zeta > 1/(\beta - 1)$, we have $n(1 - F_n(n^{\zeta})) = 0$, since $1 - F_n(n^{\zeta}) \leqslant \gamma_2 n^{-\zeta(\beta-1)} = o(n^{-1})$. So it is natural to assume that $\zeta \leqslant 1/(\beta - 1)$. Recall finally that in the Chung-Lu model [15] the maximum weight is $O(n^{1/2})$.

2.3 Results

The main theorem of this paper regards the random infection of the whole of $[n]$. We determine explicitly a critical function which we denote by $a_c(n)$ such that when we infect randomly $a(n)$ vertices in $[n]$, then the following threshold phenomenon occurs. If $a(n) \ll a_c(n)$, then a.a.s. the infection spreads no further than $\mathcal{A}_0$, but when $a(n) \gg a_c(n)$, then at least εn vertices become eventually infected, for some $\varepsilon > 0$. We remark that $a_c(n) = o(n)$.

Theorem 1. *For any* $\beta \in (2, 3)$ *and any integer* $r \geqslant 2$*, we let*

$$a_c(n) = n^{\frac{r(1-\zeta)+\zeta(\beta-1)-1}{r}} \tag{2}$$

for all $n \in \mathbb{N}$*. Let* $a : \mathbb{N} \to \mathbb{N}$ *be a function such that* $a(n) \to \infty$*, as* $n \to \infty$*, but* $a(n) = o(n)$*. Let also* $\frac{r-1}{2r-\beta+1} < \zeta \leqslant \frac{1}{\beta-1}$*. If we initially infect randomly* $a(n)$ *vertices in* $[n]$*, then the following holds:*

– *if* $a(n) \ll a_c(n)$*, then a.a.s.* $\mathcal{A}_f = \mathcal{A}_0$*;*
– *if* $a(n) \gg a_c(n)$*, then there exists* $\varepsilon > 0$ *such that a.a.s.* $|\mathcal{A}_f| > \varepsilon n$*.*

Note that the above theorem implies that when the maximum weight of the sequence is $n^{1/(\beta-1)}$, then the threshold function becomes equal to $n^{\frac{\beta-2}{\beta-1}}$ and does not depend on r.

The second theorem has to do with the targeted infection of $a(n)$ vertices where $a(n) \to \infty$, as $n \to \infty$. Let $f : \mathbb{N} \to \mathbb{R}^+$ be a function. We define the *f-kernel* to be

$$\mathcal{K}_f := \{i \in [n] \ : \ w_i \geqslant f(n)\}.$$

We will denote by $CL[\mathcal{K}_f]$ the subgraph of $CL(\mathbf{w})$ that is induced by the vertices of $\mathcal{K}_f$. We show that there exists a function f such that if we infect randomly $a(n)$ vertices of $\mathcal{K}_f$, then this is sufficient to infect almost the whole of the C-kernel, for some constant $C > 0$, with high probability. In other words, the gist of this theorem is that there is a specific part of the random graph of size $o(n)$ such that if the initially infected vertices belong to it, then this is enough to spread the infection to a positive fraction of the vertices.

Theorem 2. *Let* $a : \mathbb{N} \to \mathbb{N}$ *be a function such that* $a(n) \to \infty$*, as* $n \to \infty$*, but* $a(n) = o(n)$*. Assume also* $\frac{r-1}{2r-\beta+1} < \zeta \leqslant \frac{1}{\beta-1}$*. If* $\beta \in (2,3)$*, then there exists an* $\varepsilon_0 = \varepsilon_0(\beta, \gamma_1, \gamma_2)$ *such that for any positive* $\varepsilon < \varepsilon_0$ *there exists a constant* $C = C(\gamma_1, \gamma_2, \beta, \varepsilon, r) > 0$ *and a function* $f : \mathbb{N} \to \mathbb{R}^+$ *such that* $f(n) \to \infty$ *as* $n \to \infty$ *but* $f(n) \ll n^{\zeta}$ *satisfying the following. If we infect randomly* $a(n)$ *vertices in* $\mathcal{K}_f$*, then at least* $(1-\varepsilon)|\mathcal{K}_C|$ *vertices in* $\mathcal{K}_C$ *become infected a.a.s.*

In both theorems, the sequence of probability spaces we consider are the product spaces of the random graph together with the random choice of $\mathcal{A}_0$.

We finish this section, by stating the result of [2] concerning bootstrap percolation in the case of power-law random graphs with exponent $\beta > 3$. (Note that the result in [2] is stated for random graphs with given vertex degrees constructed through the configuration model.) We assume that at time zero each node becomes infected with probability α independently of all the other vertices. Then if p_k denotes the fraction of nodes with degree k and $p_k \propto k^{-\beta}$ for $\beta > 3$, the final fraction of infected nodes satisfies

$$\frac{|\mathcal{A}_f|}{n} \xrightarrow{p} 1 - (1-\alpha) \sum_k p_k \mathbb{P}(\mathsf{Bin}(k, 1-y^*) < r),$$

where y^* is the largest solution in $[0,1]$ to the following fixed point equation

$$y^2 \sum_k k p_k = (1-\alpha) y \sum_k k p_k \mathbb{P}(\mathsf{Bin}(k-1, 1-y) < r).$$

Our results imply that the two regimes $2 < \beta < 3$ and $\beta > 3$ have completely different behaviors.

3 Proof of Theorem 1

In this section we present a sketch of the proof of Theorem 1.

3.1 Subcritical Case

We will use a first moment argument to show that if $a(n) = o(a_c(n))$, then a.a.s. there are no vertices outside $\mathcal{A}_0$ that have at least r neighbours in $\mathcal{A}_0$ and, therefore, the bootstrap percolation process does not actually evolve. Here we assume that initially each vertex becomes infected with probability $a(n)/n$, independently of every other vertex.

For every vertex $i \in [n]$, we define an indicator random variable X_i which is 1 precisely when vertex i has at least r neighbours in $\mathcal{A}_0$. Let $X = \sum_{i\in[n]} X_i$. Our aim is to show that $\mathbb{E}[X] = o(1)$, thus implying that a.a.s. $X = 0$.

For $i \in [n]$ let $p_i = \mathbb{E}[X_i] = \mathbb{P}[X_i = 1]$. We will first give an upper bound on p_i and, thereafter, the linearity of the expected value will conclude our statement.

Lemma 1. *For all integers $r \geqslant 2$ and all $i \in [n]$, we have*

$$p_i \leqslant \left(\frac{ew_i a(n)}{rn}\right)^r.$$

From this, we can use the linearity of the expected value to deduce an upper bound on $\mathbb{E}[X]$. We have

$$\mathbb{E}[X] = \sum_{i\in[n]} p_i \leqslant \sum_{i\in[n]} \left(\frac{ew_i a(n)}{rn}\right)^r = o\left(\left(\frac{a_c(n)}{n}\right)^r\right) \sum_{i\in[n]} w_i^r. \quad (3)$$

We now need to give an estimate on $\sum_{i\in[n]} w_i^r$.

Claim. For all integers $r \geqslant 2$ and for $\beta \in (2,3)$ we have

$$\sum_{i\in[n]} w_i^r = \Theta\left(n^{1+\zeta(r-\beta+1)}\right).$$

Substituting this bound into the right-hand side of (3), we obtain:

$$\mathbb{E}[X] = o\left(\frac{n^{r(1-\zeta)+\zeta(\beta-1)-1}}{n^r} \, n^{1+\zeta(r-\beta+1)}\right).$$

But

$$r(1-\zeta) + \zeta(\beta-1) - 1 - r + 1 + \zeta(r-\beta+1) = 0,$$

thus implying that $\mathbb{E}[X] = o(1)$.

3.2 Supercritical Case

We begin with stating a recent result due to Janson, Łuczak, Turova and Vallier [22] regarding the evolution of bootstrap percolation processes on Erdős-Rényi random graphs, as these will be needed in our proofs. These results regard the binomial model $G(N,p)$ introduced by Gilbert [19] and subsequently became a major part of the theory of random graphs (see [10] or [21]). Here N is a natural number and p is a real number that belongs to $[0,1]$. We consider the set $[N] =: \{1,\ldots,N\}$ and create a random graph on the set $[N]$, including each pair $\{i,j\}$, where $i \neq j \in [N]$, independently with probability p. The following theorem from [22] considers the bootstrap percolation process on $G(N,p)$, when p as a function of N does not decay too quickly.

Theorem 3 (Theorem 5.8 [22]). *Let $r \geqslant 2$ and assume that initially a uniformly random subset of $[N]$ that has size $a(N)$ becomes infected. If $p \gg N^{-1/r}$ and $a(N) \geqslant r$, then a.a.s. $|\mathcal{A}_f| = N$.*

Now we proceed with the proof of Theorem 1. In this part of the proof, we shall be assuming that $a_c(n) = o(a(n))$. Additionally, we shall assume that the initially infected set is the set of the $a(n)$ vertices of smallest weight.

We will show first that there exists a function $f : \mathbb{N} \to \mathbb{R}^+$ such that $f(n) \to \infty$ as $n \to \infty$ but $f(n) = o(n^\zeta)$ for which a.a.s. $\mathcal{K}_f$ will become completely infected. This is where we use Theorem 3. More precisely, the subgraph of $CL(\mathbf{w})$ that is induced by the vertices of $\mathcal{K}_f$, which we denote by $CL[K_f]$, stochastically contains $G(N_f,p_f)$, where $N_f = |\mathcal{K}_f|$ and p_f is a lower bound on the probability that two vertices in $\mathcal{K}_f$ are adjacent – essentially p_f is equal to $\min\{f^2(n)/W_{[n]},1\}$. That is, one can construct a probability space that accommodates both $CL(\mathcal{K}_f)$ and $G(N_f,p_f)$, on the same vertex set and with the correct distributions, in such a way that always the latter is a subgraph of the former.

We then show that any given vertex in $\mathcal{K}_f$ has at least r neighbours in $\mathcal{A}_0$ with some probability p_{Inf} which we determine later in (4). In other words, each vertex in $\mathcal{K}_f$ becomes infected in one round with probability p_{Inf} independently of every other vertex. Hence, as we may consider $G(N_f,p_f)$ as a subgraph of $CL[\mathcal{K}_f]$ on the same vertex set, we deduce that the final set of infected vertices in $\mathcal{K}_f$ is bounded from below by the size of the final set of infected vertices in a bootstrap percolation process on $G(N_f,p_f)$, assuming that the set of initially infected vertices is the set of vertices which have at least r neighbours in $\mathcal{A}_0$. We will show that p_{Inf}, N_f and p_f satisfy the premises of Theorem 3, whereby we will deduce that in fact $\mathcal{K}_f$ becomes completely infected a.a.s. Thereafter, we use the following proposition, whose proof is rather lengthy and technical and, for this reason, we omit it. We consider a bootstrap percolation process on $CL(\mathbf{w})$ where the initially infected set is a large subset of $\mathcal{K}_f$.

Proposition 1. *Let $r \geqslant 2$ and let $f : \mathbb{N} \to \mathbb{R}^+$ be a function such that $f(n) \to \infty$ as $n \to \infty$ but $f(n) = o(n^\zeta)$. Then there exists an $\varepsilon_0 = \varepsilon_0(\beta,\gamma_1,\gamma_2) > 0$ such that for any positive $\varepsilon < \varepsilon_0$ there exists $C = C(\gamma_1,\gamma_2,\beta,\varepsilon,r) > 0$ for which the following holds. If $(1-\varepsilon)|\mathcal{K}_f|$ vertices of $\mathcal{K}_f$ have been infected, then a.a.s. at least $(1-\varepsilon)|\mathcal{K}_C|$ vertices of $\mathcal{K}_C$ become infected.*

We deduce by above proposition that there exists a real number $C > 0$ such that with high probability $\mathcal{K}_C$ will be almost completely infected. This and Definition 2 imply that there exists an $\varepsilon > 0$ such that a.a.s. at least εn vertices become infected.

Spreading the Infection to a Positive Fraction of the Vertices. We begin with determining the function f. To this end, we need to bound from below the probability that an arbitrary vertex in $\mathcal{K}_f$ becomes infected. In fact, we shall bound from below the probability that an arbitrary vertex in $\mathcal{K}_f$ will become infected already in the first round. Note that this amounts to bounding the probability that such a vertex has at least r neighbours in $\mathcal{A}_0$. Therefore, this forms a collection of independent events which is equivalent to the random independent infection of the vertices of $\mathcal{K}_f$ with probability equal to the derived lower bound. Recall that the random graph induced on $\mathcal{K}_f$ stochastically contains an Erdős-Rényi random graph with the appropriate parameters. This observation allows us to determine f. To be more specific, if the probability that any given vertex in $\mathcal{K}_f$ exceeds the complete infection threshold of this Erdős-Rényi random graph and the premises of Theorem 3 is satisfied, then a.a.s. $\mathcal{K}_f$ eventually becomes completely infected.

Under the assumption that $\mathcal{A}_0$ consists of the $a(n)$ vertices of smallest weight, we will bound from below the probability a vertex $v \in \mathcal{K}_f$ has at least r neighbours in $\mathcal{A}_0$. We denote the degree of v in $\mathcal{A}_0$ by $d_{\mathcal{A}_0}(v)$ and note that this random variable is equal to $\sum_{i\in\mathcal{A}_0} \mathsf{Be}\left(\frac{w_v w_i}{W_{[n]}}\right)$, where the summands are independent Bernoulli distributed random variables. Note also that for all n and for all $i \in [n]$ we have $w_i \geqslant x_0$. Thus, we can deduce the following (parts of it hold for n sufficiently large)

$$\begin{aligned}
\mathbb{P}\left[\sum_{i\in\mathcal{A}_0} \mathsf{Be}\left(\frac{w_v w_i}{W_{[n]}}\right) \geqslant r\right] &\geqslant \mathbb{P}\left[\sum_{i\in\mathcal{A}_0} \mathsf{Be}\left(\frac{w_v x_0}{W_{[n]}}\right) \geqslant r\right] \\
&= \mathbb{P}\left[\mathsf{Bin}\left(a(n), \frac{w_v x_0}{W_{[n]}}\right) \geqslant r\right] \\
&\geqslant \binom{a(n)}{r} \left(\frac{w_v x_0}{W_{[n]}}\right)^r \left(1 - \frac{w_v x_0}{W_{[n]}}\right)^{a(n)-r} \\
&\geqslant \frac{a(n)^r}{1.5\ r!} \left(\frac{f(n) x_0}{W_{[n]}}\right)^r \left(1 - \frac{f(n) x_0}{W_{[n]}}\right)^{a(n)-r} .
\end{aligned}$$

Thus, assuming that $a(n)f(n) = o(n)$ we have

$$\left(1 - \frac{f(n)x_0}{W_{[n]}}\right)^{a(n)-r} = 1 - o(1).$$

Therefore, for n sufficiently large

$$\mathbb{P}\left[\sum_{i\in\mathcal{A}_0} \mathsf{Be}\left(\frac{w_v w_i}{W_{[n]}}\right) \geqslant r\right] \geqslant \frac{1}{2r!} \left(\frac{a(n)f(n)x_0}{W_{[n]}}\right)^r =: p_{Inf}. \tag{4}$$

Thus, every vertex of $\mathcal{K}_f$ becomes infected during the first round with probability at least p_{Inf}, independently of every other vertex in $\mathcal{K}_f$.

Recall that $\frac{2r-\beta+1}{r-1} \leqslant \zeta \leqslant \frac{1}{\beta-1}$ and $a_c(n) = n^{\frac{r(1-\zeta)+\zeta(\beta-1)-1}{r}}$. Let us assume that $a(n) = \omega(n) n^{\frac{r(1-\zeta)+\zeta(\beta-1)-1}{r}}$, where $\omega : \mathbb{N} \to \mathbb{R}^+$ is some increasing function that grows slower than any polynomial. Setting $f = f(n) = \frac{n^\zeta}{\omega^{1+1/r}(n)}$, we will consider $CL[\mathcal{K}_f]$. Before doing so, we will verify the assumption that $a(n)f(n) = o(n)$. Indeed, we have

$$a(n)f(n) = \frac{1}{\omega^{1/r}(n)}\; n^{\frac{r(1-\zeta)+\zeta(\beta-1)-1}{r}+\zeta}.$$

But

$$\begin{aligned}\frac{r(1-\zeta)+\zeta(\beta-1)-1}{r} + \zeta &= \frac{r(1-\zeta)+\zeta(\beta-1)-1+r\zeta}{r} \\ &= 1 + \frac{\zeta(\beta-1)-1}{r} \leqslant 1,\end{aligned}$$

since $\zeta \leqslant 1/(\beta-1)$, whereby $a(n)f(n) \leqslant \frac{n}{\omega^{1/r}(n)} = o(n)$.

Now, note that if $\zeta > \frac{1}{2}$, then $CL[\mathcal{K}_f]$ is the complete graph on $|\mathcal{K}_f|$ vertices. However, when $\zeta \leqslant \frac{1}{2}$, then $CL[\mathcal{K}_f]$ stochastically contains $G(N_f, p_f)$, where $N_f = |\mathcal{K}_f|$ and $p_f = \frac{f^2(n)}{W_{[n]}}$. We will treat these two cases separately.

Case I: $\frac{1}{2} < \zeta \leqslant \frac{1}{\beta-1}$.

In this case, as $CL[\mathcal{K}_f]$ is the complete graph, it suffices to show that with high probability at least r vertices of $\mathcal{K}_f$ become infected already at the first round. In fact, we will show that the expected number of vertices of $\mathcal{K}_f$ that become infected during the first round tends to infinity as n grows. Note that this number is equal to $N_f p_{Inf}$. Thus, once we show that $N_f p_{Inf} \to \infty$, as $n \to \infty$, then Chebyschev's inequality or a standard Chernoff bound can show that with probability $1-o(1)$, there are at least r infected vertices in $\mathcal{K}_f$ and, thereafter, the whole of $\mathcal{K}_f$ becomes infected in one round.

By Definition 2 we have

$$N_f = |\mathcal{K}_f| = \Omega\left(n\left(\frac{\omega(n)}{n^\zeta}\right)^{\beta-1}\right),$$

and by (4) we have

$$p_{Inf} = \Theta\left(\frac{1}{\omega(n)}\;\left(\frac{n^{\frac{r(1-\zeta)+\zeta(\beta-1)-1}{r}} \cdot n^\zeta}{n}\right)^r\right) = \Theta\left(\frac{n^{\zeta(\beta-1)-1}}{\omega(n)}\right).$$

Hence

$$N_f p_{Inf} = \Omega\left(\omega^{\beta-2}(n)\right).$$

Case II: $\frac{r-1}{2r-\beta+1} < \zeta \leqslant \frac{1}{2}$.

As we mentioned above, $CL[\mathcal{K}_f]$ stochastically contains $G(N_f, p_f)$, where $p_f = \frac{f^2(n)}{W_{[n]}}$, as $\zeta \leqslant \frac{1}{2}$. We will show that here $N_f p_f^r \to \infty$ as $n \to \infty$ and by Theorem 3 we deduce that $\mathcal{K}_f$ becomes completely infected with probability $1-o(1)$. We have

$$N_f p_f^r = \Theta\left(\omega^{\beta-1}(n) n^{1-\zeta(\beta-1)} \frac{n^{2\zeta r}}{\omega^{2r+2}(n) n^r}\right). \quad (5)$$

and the expression on the right-hand side is

$$\omega^{-(2r-\beta+3)}(n) n^{-(r-1)+\zeta(2r-\beta+1)} \to \infty,$$

by our assumption on ζ.

For each one of the above cases, Proposition 1 implies that for any real $\varepsilon > 0$ that is small enough there exists a real number $C = C(\gamma_1, \gamma_2, \beta, \varepsilon) > 0$ such that a.a.s. at least $(1-\varepsilon)|\mathcal{K}_C|$ vertices of $\mathcal{K}_C$ become infected. But we have $|\mathcal{K}_C| = \Theta(n)$ and the second part of Theorem 1 follows.

4 Conclusion

In this paper, we analyse the evolution of a bootstrap percolation process in a class of inhomogeneous random graphs which exhibits a power law degree distribution with exponent β between 2 and 3. The main result of this work is that a sublinear initially infected set is enough to spread the infection to a linear fraction of vertices of the random graph. We further explore the role of hub vertices of the random graph and demonstrate their function in the evolution of the process.

Acknowledgment. We would like to thank Rob Morris for drawing our attention to an oversight in an earlier version of this paper.

References

1. Adler, J., Lev, U.: Bootstrap percolation: visualizations and applications. Brazilian Journal of Physics 33(3), 641–644 (2003)
2. Amini, H.: Bootstrap percolation and diffusion in random graphs with given vertex degrees. Electronic Journal of Combinatorics 17, R25 (2010)
3. Amini, H.: Bootstrap percolation in living neural networks. Journal of Statistcal Physics 141, 459–475 (2010)
4. Amini, H., Cont, R., Minca, A.: Resilience to contagion in financial networks (2011), preprint `http://ssrn.com/abstract=1865997`
5. Balogh, J., Bollobás, B.: Bootstrap percolation on the hypercube. Probability Theory and Related Fields 134(4), 624–648 (2006)
6. Balogh, J., Bollobás, B., Duminil-Copin, H., Morris, R.: The sharp threshold for bootstrap percolation in all dimensions. Trans. Amer. Math. Soc. 364, 2667–2701 (2012)
7. Balogh, J., Bollobás, B., Morris, R.: Bootstrap percolation in three dimensions. Annals of Probability 37, 1329–1380 (2009)
8. Balogh, J., Peres, Y., Pete, G.: Bootstrap percolation on infinite trees and non-amenable groups. Combinatorics, Probability and Computing 15(5), 715–730 (2006)
9. Balogh, J., Pittel, B.G.: Bootstrap percolation on the random regular graph. Random Structures & Algorithms 30(1-2), 257–286 (2007)

10. Bollobás, B.: Random Graphs. Cambridge studies in advanced mathematics, 2nd edn. Cambridge University Press (2001)
11. Bollobás, B., Janson, S., Riordan, O.: The phase transition in inhomogeneous random graphs. Random Structures & Algorithms 31(1), 3–122 (2007)
12. Cerf, R., Manzo, F.: The threshold regime of finite volume bootstrap percolation. Stochastic Processes and their Applications 101(1), 69–82 (2002)
13. Chalupa, J., Leath, P.L., Reich, G.R.: Bootstrap percolation on a Bethe lattice. Journal of Physics C: Solid State Physics 12, L31–L35 (1979)
14. Chung, F., Lu, L.: Connected components in random graphs with given expected degree sequences. Annals of Combinatorics 6, 125–145 (2002)
15. Chung, F., Lu, L.: The average distance in a random graph with given expected degrees. Internet Mathematics 1(1), 91–113 (2003)
16. Chung, F., Lu, L., Vu, V.: The spectra of random graphs with given expected degrees. Internet Mathematics 1(3), 257–275 (2004)
17. Faloutsos, M., Faloutsos, P., Faloutsos, C.: On power-law relationships of the internet topology. In: SIGCOMM 1999: Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, pp. 251–262 (1999)
18. Fontes, L., Schonmann, R.: Bootstrap percolation on homogeneous trees has 2 phase transitions. Journal of Statistical Physics 132, 839–861 (2008)
19. Gilbert, E.N.: Random graphs. Annals of Mathematical Statistics 30, 1141–1144 (1959)
20. Holroyd, A.E.: Sharp metastability threshold for two-dimensional bootstrap percolation. Probability Theory and Related Fields 125(2), 195–224 (2003)
21. Janson, S., Łuczak, T., Ruciński, A.: Random graphs. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley Interscience, New York (2000)
22. Janson, S., Łuczak, T., Turova, T., Vallier, T.: Bootstrap percolation on the random graph $G_{n,p}$. To Appear in The Annals of Applied Probability (2010), http://arxiv.org/abs/1012.3535
23. Kumar, R., Raghavan, P., Rajagopalan, S., Tomkins, A.: Extracting large scale knowledge bases from the web. In: Proceedings of the 25th VLDB Conference, pp. 639–650 (1999)
24. Sabhapandit, S., Dhar, D., Shukla, P.: Hysteresis in the random-field Ising model and bootstrap percolation. Physical Review Letters 88(19), 197202 (2002)
25. Söderberg, B.: General formalism for inhomogeneous random graphs. Physical Review E 66, 066121 (2002)
26. Tlusty, T., Eckmann, J.P.: Remarks on bootstrap percolation in metric networks. Journal of Physics A: Mathematical and Theoretical 42, 205004 (2009)
27. Toninelli, C., Biroli, G., Fisher, D.S.: Jamming percolation and glass transitions in lattice models. Physical Review Letters 96(3), 035702 (2006)
28. van der Hofstad, R.: Random Graphs and Complex Networks (2011) (book in preparation), http://www.win.tue.nl/rhofstad/NotesRGCN2011.pdf
29. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: KDD 2003: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137–146. ACM, New York (2003)
30. Chen, N.: On the approximability of influence in social networks. SIAM Journal on Discrete Mathematics 23(3), 1400–1415 (2009)
31. Kleinberg, J.: Cascading behavior in networks: algorithmic and economic issues. In: Algorithmic Game Theory, pp. 613–632. Cambridge University Press, Cambridge (2007)