Bounds for critical values of the Bak-Sneppen model on transitive graphs

Alexis Gillett, Ronald Meester and Misja Nuyens
Vrije Universiteit Amsterdam, The Netherlands

Abstract

We study the Bak-Sneppen model on locally finite transitive graphs G, in particular on $\mathbb{Z}^d$ and on T_Δ, the regular tree with common degree Δ. We show that the avalanches of the Bak-Sneppen model dominate independent site percolation, in a sense to be made precise. Together with the fact that avalanches of the Bak-Sneppen model are dominated by a simple branching process, this yields upper and lower bounds for the critical value $p_c^{BS}(G)$ of the Bak-Sneppen model. Our main results state that $\frac{1}{\Delta+1} \le p_c^{BS}(T_\Delta) \le \frac{1}{\Delta-1}$, and that $\frac{1}{2d+1} \le p_c^{BS}(\mathbb{Z}^d) \le \frac{1}{2d} + \frac{1}{(2d)^2} + O\big(d^{-3}\big)$, as $d \to \infty$.

Keywords: Bak-Sneppen model, critical values, coupling, site percolation, branching process
2000 MSC: 60K35, 82B43 (primary); 60J80, 82C22 (secondary)

1 Introduction and main results

The Bak-Sneppen model was originally introduced as a simple model of evolution by Per Bak and Kim Sneppen [2]. Their original model can be defined as follows. There are N species arranged on a circle, each of which has been assigned a random *fitness*. The fitnesses are independent and uniformly distributed on $(0, 1)$. At each discrete time step the system is updated by locating the lowest fitness and replacing this fitness, and those of its two neighbours, by independent and uniform $(0, 1)$ random variables.

It is not particularly significant that the underlying graph of the model is the circle, or $\mathbb{Z}$ in the thermodynamic limit. Bak-Sneppen models can be defined on a wide range of graphs using the same update rule as above. In this more general situation, the fitnesses of the vertex with minimal fitness and its neighbours are updated. In particular, Bak-Sneppen *avalanches* (defined below) can be defined on any locally finite graph. Unlike particle systems such as percolation or the contact process, the Bak-Sneppen model has no tuning parameter. Therefore, it has been described as exhibiting self-organised critical behaviour, see [8] for a discussion.

One of the ways to analyse Bak-Sneppen models is to break them down into a series of avalanches. An avalanche from a *threshold* p, referred to as a p-*avalanche*, is said to occur between times s and $s+t$ if at time s all the fitnesses are equal to or greater than p with at most one vertex where equality holds, and time $s+t$ is the next time after s at which this occurs. In the literature alternative types of avalanches have been proposed, see [6, 7]. The definition given here agrees with the most commonly used notion of an avalanche and was introduced by Bak and Sneppen [2]. A summary of avalanche properties is given below. For a more thorough coverage readers are directed to Meester and Znamenski [9, 10].

In this paper, we look only at *transitive* graphs. The behaviour of an avalanche on a transitive graph is independent of its origin; an avalanche started at vertex i behaves the same as an avalanche started at the origin. When analysing avalanches on transitive graphs, it is therefore natural to talk about a typical p-avalanche. We remark that our results could be formulated in the general locally finite setting, without the assumption of transitivity. However, this would introduce a number of technicalities that obscure the main lines of reasoning for little practical gain.

A Bak-Sneppen model on an *infinite* graph is not well-defined, since it is not possible to locate the minimal fitness when there are infinitely many vertices to check. However, an avalanche on a locally finite graph doesn't have this problem, as the minimal fitness during an avalanche always appears among the vertices already updated by the avalanche. The ability to look directly at infinite graphs is very desirable, because the most interesting behaviour of the Bak-Sneppen model is observed in the limit as the number of vertices in the graph tends to infinity.

To analyse avalanches, some definitions are needed. The set of vertices updated by an avalanche is referred to as its range set, with its *range* being the cardinality of this range set. The *duration* of the avalanche is the number of updates during the avalanche. It can be seen that the events that an avalanche has infinite range and infinite duration differ on a set of null measure, see [10], so we can talk about an infinite avalanche with no danger of confusion. Letting $r_G^{BS}(p)$ denote the range of a p-avalanche on a transitive graph G, we define the *critical value* of the Bak-Sneppen model as

$$p_c^{BS}(G) = \inf\{p > 0 : \mathbb{P}(r_G^{BS}(p) = \infty) > 0\}.$$

It is also possible to define a critical value based on the expected range. It is believed, but unproven, that for the Bak-Sneppen model these two definitions yield the same critical value. Although throughout this paper we consider the critical value of the Bak-Sneppen model given above, our results also give bounds for the other definition of the critical value. Numerical simulations [2] suggest that the stationary marginal fitness distributions for the Bak-Sneppen model on N sites tends to a uniform distribution on $(p_c^{BS}(\mathbb{Z}^d), 1)$, as $N \to \infty$. Furthermore, it has been proved that if there is a unique phase transition, i.e., if the critical values defined above are equal, then this is indeed the case, see [10].

The approach of this paper is to compare Bak-Sneppen avalanches with two other well-known processes, namely branching processes and independent site percolation. A simple comparison with branching processes gives a lower bound on the critical value, whereas a much more complicated comparison with site percolation gives an upper bound. To warm up, we first give the (easy) lower bound.

Proposition 1.1 *On any locally finite transitive graph G with common vertex degree Δ, we have*

$$p_c^{BS}(G) \geq \frac{1}{\Delta + 1}.$$

Proof: At every update of the system, we draw $\Delta + 1$ independent uniform $(0, 1)$ random variables to get the new fitnesses of the vertex with minimal fitness, and of its Δ neighbours. Each of these $\Delta + 1$ new fitnesses is below the threshold p with probability p, independent of each other. This induces a coupling with a simple branching process with binomial $(\Delta + 1, p)$

offspring distribution, where every active vertex in the Bak-Sneppen avalanche is represented by at least one particle in the branching process. Hence, if the branching process dies out, then so does the Bak-Sneppen avalanche. Therefore the critical value of the Bak-Sneppen avalanche can be no smaller than the critical value of the branching process. □

The main result of this paper is the following upper bound for the critical value $p_c^{BS}(G)$ of the Bak-Sneppen model on a locally finite transitive graph G. The critical value for independent site percolation on G is denoted by $p_c^{site}(G)$.

Theorem 1.2 *On any locally finite transitive graph G we have*

$$p_c^{BS}(G) \leq p_c^{site}(G).$$

Since the critical value of site percolation on T_Δ, the regular tree with common degree Δ, equals $1/(\Delta - 1)$, the following corollary holds.

Corollary 1.3 *The critical value of the Bak-Sneppen model on a regular tree with common degree Δ satisfies*

$$\frac{1}{\Delta+1} \leq p_c^{BS}(T_\Delta) \leq \frac{1}{\Delta-1}.$$

Applying the expansion for the critical value of site percolation on $\mathbb{Z}^d$ given by Hara and Slade [5], we have the following corollary.

Corollary 1.4 *The critical value of the Bak-Sneppen model on $\mathbb{Z}^d$ satisfies*

$$\frac{1}{2d+1} \leq p_c^{BS}(\mathbb{Z}^d) \leq \frac{1}{2d} + \frac{1}{(2d)^2} + O\big(d^{-3}\big), \qquad d \to \infty.$$

The paper is organised as follows. In Section 2 we take some preliminary steps by describing an alternative way of constructing a Bak-Sneppen avalanche. Section 3 uses this construction to couple the Bak-Sneppen avalanche and another stochastic process. The proof that the critical value of the Bak-Sneppen avalanche is larger than that of the coupled process is given in Section 4. The proof of Theorem 1.2 is completed in Section 5 where we show that the coupled stochastic process is in fact site percolation.

2 An alternative construction of the Bak-Sneppen model

In the introduction the Bak-Sneppen model was defined in its original format and then generalised to locally finite graphs. However, for our purposes it is more convenient to work with an alternative construction of the Bak-Sneppen model. We call this new construction the *forgetful* Bak-Sneppen model, as the exact fitness values will be no longer fixed (or remembered). This idea borrows heavily from the 'locking thresholds representation' in [9], and was used in a much simpler form in [3]. The forgetful Bak-Sneppen model is defined below and then argued to be equivalent to the normal Bak-Sneppen model, in the sense that at all times, the fitnesses have the same distributions.

Consider a Bak-Sneppen model on a finite transitive graph G with N vertices. To start with, N independent uniform $(0,1)$ random variables are drawn, one for each vertex. After that we proceed as follows.

1. The minimum of the fitnesses is found and fixed.
2. All the other fitness values are discarded, and replaced by the conditional *distribution* of these fitnesses, given that they are larger than the observed minimum fitness.
3. The minimal vertex and its neighbours have their value/distributions replaced by uniform (0,1) distributions. (So now all vertices have some distribution associated with them.)
4. Now N new independent random variables are drawn according to the appropriate fitness distributions, and we go back to the first step.

Let the fitnesses drawn in the last step (at some moment in time) be denoted by $X = (X_1, X_2, \ldots, X_N)$. It is easy to see, using induction, that X has the correct distribution.

Note that all conditional fitness distributions have one convenient property, namely that they are all uniform distributions. Indeed, suppose that a random variable Y has a uniform $(y,1)$ distribution, denoted by F_y. If we condition on $Y > z$, then Y has distribution $F_{y \vee z}$, where $y \vee z = \max\{y, z\}$. All our fitnesses initially have uniform (0,1) distributions. Two things can change these distributions. They can be reset to F_0 by being updated, or they can be conditioned to be bigger than some given value; in both cases they remain uniform.

The above construction gives a Bak-Sneppen type model on a finite graph, but this is easily extended to a Bak-Sneppen avalanche on a locally finite graph. The only difference is that initially we only assign $\Delta + 1$ fitnesses to the origin and its Δ neighbours. The remaining vertices are considered to have no fitness until they are updated by the avalanche. The avalanche ends when all the fitnesses within the avalanche are above the threshold, which is equivalent to saying that the minimal fitness is above the threshold. So we can use the forgetful method to generate avalanches, which stop when the minimum is found to be above the threshold.

3 The construction of the coupling

This section is divided into three parts. To begin with, some intuition behind the result is given. This is followed by a precise description of the coupling and then for added clarity we give an example.

3.1 Intuition

We are interested in comparing the open cluster at the origin of site percolation to a Bak-Sneppen avalanche. Typically, site percolation is considered to be static, but it is also possible to build up the open cluster at the origin dynamically. This can be done as follows. To begin with, consider the origin to be open and look at its neighbours. Decide which of these vertices are open. Then look at the new neighbours of the open cluster and iterate.

The growth of both a Bak-Sneppen avalanche and the open cluster at the origin is driven by the *extremal* vertices. In a Bak-Sneppen avalanche, the extremal vertices are those vertices that are contained within the avalanche and have neighbours outside the avalanche. It is only through one of the extremal vertices being minimal that the range of the avalanche can increase. For site percolation, the extremal vertices are those having a neighbour in the open cluster at the origin, but that are themselves unknown as to be open or closed. These are exactly the vertices at the edge of the cluster that will increase the size of the cluster by being open. Since it is the extremal vertices that drive the spread of both processes, the task is to relate the two sets of extremal vertices to each other.

The major difficulty to overcome is that in the Bak-Sneppen model an extremal vertex may be updated many times by neighbouring activity before being minimal itself, whereas in site percolation a vertex is either open or closed. So in the Bak-Sneppen model it is possible that a previously active extremal vertex is never minimal, having been made inactive by a subsequent neighbouring update. Conversely, it is also possible that an originally inactive vertex can be made active. This means that coupling the two models in the natural (direct) manner is not useful, since an open vertex in the site percolation may later find its equivalent in the Bak-Sneppen model inactive (and vice versa).

However, if a vertex's fitness is not minimal, then its conditional distribution based on this information is stochastically larger than its original uniform $(0, 1)$ distribution. So if a vertex is updated by a neighbour being minimal, this makes its fitness stochastically smaller, making the vertex more likely to be active and therefore, intuitively at least, the avalanche is more likely to continue. This means that on average the interference from the non-extremal vertices of the Bak-Sneppen model on the extremal vertices should be beneficial to the spread of the avalanche.

3.2 The coupling

We now describe the construction of a process that we will refer to as the *coupled process*. First fix an avalanche threshold p.

We start with two copies of the graph G, denoted G_B and G_C respectively. Initially we assign the value 0 to the origin of both graphs, and all the Δ neighbours of the origin of both graphs get distribution F_0. On G_C, we define the *extremal set* $\mathcal{E}$ as the set of all points that have been assigned a distribution, but not (yet) an exact value.

We update the Bak-Sneppen avalanche on G_B according to the aforementioned (forgetful) construction, i.e., we generate new fitnesses, fix the minimal value and then update the distributions of the other fitnesses accordingly. In the coupled process, only the vertices contained in $\mathcal{E}$ are updated. This goes as follows. Consider a vertex $v_C \in \mathcal{E}$ with G_B-counterpart v_B. Let F_z and F_y be their respective fitness distributions. We realise the fitnesses of the vertices in G_B, and in particular realise the fitness of v_B with an independent uniform (0,1) distributed random variable U via $y+(1-y)U$. Let M be the minimal fitness in G_B. As long as the minimal vertex in the avalanche is active, i.e., $M \leq p$, we have the following two options, with corresponding update rules. Below these are illustrated by an explicit example.

1. The fitness of v_B is not minimal.

   We update the conditional distribution of v_C using the extra information that the fitness of v_B must be bigger than M. Since the fitness of v_B is not minimal, we have $y+(1-y)U > M$, and hence $U > (M-y)^+/(1-y)$. The conditional distribution of v_C after the update is $F_{\hat{z}}$, where

   $$\hat{z} = z + (1-z)\frac{(M-y)^+}{(1-y)}.$$

2. The fitness of v_B is minimal, so it has value M.

   It follows that $y+(1-y)U = M$. The fitness of v_C is now **fixed** at

   $$z + (1-z)U = z + (1-z)\frac{(M-y)}{(1-y)}.$$

   If this value is less than p, we say that v is *open*, remove v from $\mathcal{E}$, add the neighbours of v that have an undetermined state to $\mathcal{E}$, and give them distribution F_0. If the value of v is larger than p, then v is *closed* and removed from $\mathcal{E}$.

The final step of the construction is as follows. The first moment the minimal fitness in G_B is inactive (that is, is larger than p), the Bak-Sneppen avalanche has finished. As soon as this happens, we fix *all* the values of the vertices in $\mathcal{E}$ in a similar way to rule 2 above: if the vertex in $\mathcal{E}$ has distribution F_z and associated random variable U, then its value is $z+(1-z)U$. Below we show that as soon as the Bak-Sneppen avalanche ends, this value is at least p, and hence all the vertices in $\mathcal{E}$ will be *closed*.

3.3 An example

The update rules are illustrated by the following example, displayed in Figure 1. In this example the graph G is a tree. For illustration purposes, we show only the part of the graph where the activity takes place.

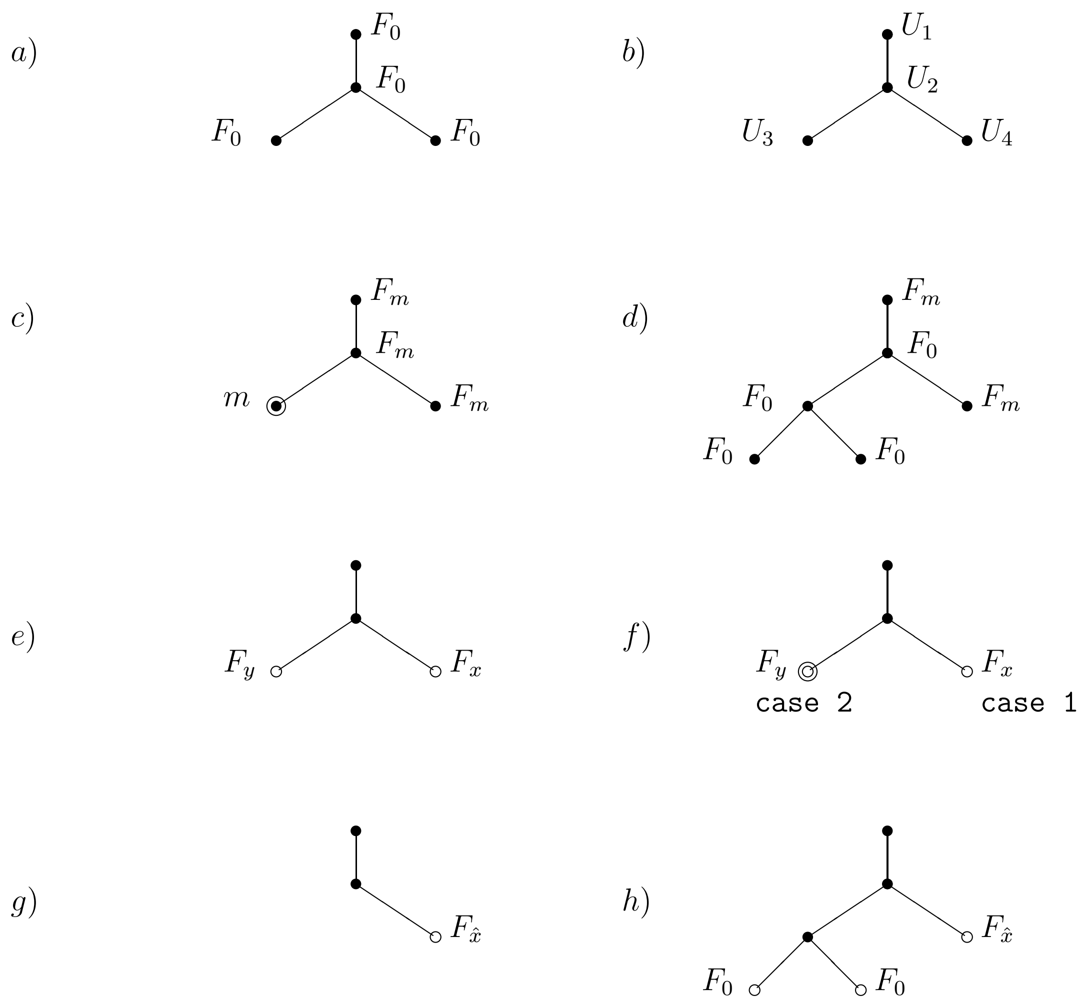

Figure 1: **A Bak-Sneppen update.** Graph a shows the state of the system before the update and graph d shows the state of the system before the next update. Graphs b and c show the intermediate steps. The forgetful Bak-Sneppen model never assumes the values given in graph b. **The same update in the coupled process.** Black points are open, white points are undetermined. Graph g gives the case when the F_y random variable is larger than the threshold p and graph h when it is smaller than p. Note that $\hat{x} = x + (1 - x)m$.

We consider the avalanche just before an update step, in the following situation, see Figure 1, graph a: all fitnesses have distribution F_0. Then in graph b, uniform $(0,1)$ random variables U_1, U_2, U_3, and U_4 are drawn and used to calculate the location and magnitude of the new minimal fitness (graph c). Finally, the new fitness distributions are determined, and we are in a situation similar to graph a, just before the next update step (graph d).

During the same update step, the coupled process evolves as follows. We only consider the vertices in the extremal set $\mathcal{E}$, see Figure 3, graph e. Before the update, the vertices have distributions F_y and F_x. Given the location of the minimum fitness in the Bak-Sneppen avalanche, the vertices in $\mathcal{E}$ are classified according to the cases 1. and 2. above (graph f). From the location and magnitude of the minimal fitness of the avalanche it follows that $U_4 \geq M$, so $\hat{x} = x + (1-x)M$. Finally, the value of the vertex that corresponds with the minimal vertex in the avalanche is fixed, according to update rule 2. Its value f is given by $f = y + (1-y)M$. Now there are two possible cases: either $f > p$, and the vertex is closed (graph g), or $f \leq p$, and the vertex is open and its undetermined neighbours receive distribution F_0 (graph h).

4 A domination principle

To show that the critical value of the coupled process can be no smaller than the critical value of the Bak-Sneppen avalanche, we use a domination argument. The propositions below show that the coupled process can finish no later than the Bak-Sneppen avalanche (so that the avalanche can be said to dominate the coupled process).

Proposition 4.1 *For every $v_C \in G_C$ and corresponding $v_B \in G_B$, at all times, the (conditional) fitness distribution of v_C is stochastically larger than the (conditional) fitness distribution of v_B.*

Proof: It should be noted that this proposition only makes sense for vertices in $\mathcal{E}$. Furthermore, it is safe to assume that the p-avalanche is still in progress, so the minimal fitness is less than p.

The proof proceeds by induction. When new vertices are added to the coupled process, they (and their equivalents in G_B) have uniform $(0,1)$ distributed fitnesses. This is by definition for the coupled process, but also holds for G_B, since vertices in G_B corresponding to new vertices added to $\mathcal{E}$ are always neighbours of the minimal vertex. This means that the statement of the proposition holds for new vertices added to $\mathcal{E}$.

To make the induction step, consider $v_C \in \mathcal{E}$ with corresponding vertex v_B in G_B, and let F_z and F_y be the fitness distributions of v_C and v_B before the update step, where $y \leq z < 1$. Let u be the realisation of the uniform $(0,1)$ random variable associated with v_C and v_B at the update step, and let m be the minimal avalanche fitness. Assume first that v_B does not have the minimal fitness. This provides information on the value of u, namely that $y + (1-y)u > m$, and hence $u > (m-y)/(1-y)$. If $y \geq m$, this information is useless: we already knew that $u > 0$, and the distributions of v_C and v_B are not updated. If $y < m$, the inequality for u does contain information, and we can calculate the corresponding inequality for the fitness of v_C:

$$z + (1-z)u > z + \frac{(1-z)(m-y)}{1-y} = m + \frac{(1-m)(z-y)}{1-y} := \hat{z}.$$

So after the update, v_B has distribution $F_{y \vee m}$ and v_C has distribution $F_{\hat{z}}$. Since $m, y < 1$ and $y \leq z$, we have $\hat{z} \geq m$. Hence $(y \vee m) \leq \hat{z}$, and the desired property holds.

Second, we consider the case that a neighbour of v_B was minimal. In that case the fitness distribution of v_B is reset to F_0, and there is nothing left to prove. □

Proposition 4.2 *At the moment that the p-avalanche ends, all vertices in $\mathcal{E}$ are closed. As a consequence, if the probability of an infinite p-avalanche is zero, then there cannot be an infinite cluster of open sites in the coupled process, almost surely.*

Proof: By Proposition 4.1, at any stage every point in $\mathcal{E}$ has a fitness that is stochastically larger than the fitness of the corresponding vertex in the avalanche. Hence, if the p-avalanche ends, then in the coupled process all neighbours in the set $\mathcal{E}$ will be closed, as their fixed values can not be smaller than those in the avalanche, which are already greater than p as the avalanche has ended. This removes all vertices from $\mathcal{E}$ and ensures that no more are added, implying that in the coupled process no more vertices will be added to the open cluster around the origin. □

5 The coupled process is independent site percolation

To complete the proof of Theorem 1.2, it remains to show that the coupled process is in fact independent site percolation.

Our first goal is to determine the distribution of the information we use to update the coupled process. More precisely, consider an arbitrary step of the forgetful Bak-Sneppen model, when there are n vertices in the avalanche range so far. We enumerate these vertices $1, \dots, n$ and suppose that all n vertices in the avalanche have just been assigned a (conditional) distribution $F_{y_1}, \dots, F_{y_n}$. (Recall that these are just uniform distributions above the respective y_i's.) We sample from this random vector, using independent uniform $(0,1)$ distributed random variables $U_1, \dots, U_n$: a sample from F_{y_i} is realised via $y_i + (1 - y_i)U_i$. We locate the minimum M, at vertex K; note that both M and K are random. This clearly implies that

$$U_K = \frac{M - y_K}{1 - y_K}.$$

The remaining values are now uniformly distributed above $\max\{y_i, M\}$ respectively, for all $i \neq K$, that is, we know that

$$U_i > \frac{(M - y_i)^+}{1 - y_i}, \qquad i \neq K.$$

Taking $U_K = (M - y_K)/(1 - y_K)$ and sampling from all the other entries (which are uniformly distributed above $\frac{(M-y_i)^+}{1-y_i}$ respectively) is just a somewhat complicated way of sampling from the original vector $(U_1, \dots, U_n)$, that is, such a sample yields independent uniform $(0,1)$ distributed entries.

Looking back to Section 3.2, it should be clear that the update rules of the coupled process use independent uniform $(0,1)$ random variables generated in the above way to update the vertices of the coupled process contained in $\mathcal{E}$. Note that this trimming of the random variables (we use $|\mathcal{E}|$ rather than n) doesn't affect their marginal distributions or dependence structure as the trimming is done independently of the values of the uniform random variables.

It is now possible to give a direct description of the construction of the coupled process. We start with the origin being open and look at the neighbours of the origin, which initially have distribution F_0. These distributions are realised as follows, using the independent uniform $(0,1)$ random variables described above. At each time step at most one value becomes fixed and the rest are given updated distributions. The fixed value corresponds to the case that $U_K = (M - y_K)/(1 - y_K)$ relates to a vertex in $\mathcal{E}$. To update the other vertices in $\mathcal{E}$, we use the information that the U_i's are distributed independently and uniformly above $\frac{(M-y_i)^+}{1-y_i}$. This means that we do not fix their actual values at that time step, but instead change their fitness distributions conditioned on this information. Once a vertex has a fixed value, it is declared open if and only if this value falls below p. Whenever a vertex is declared open, those neighbours that neither have a fixed value nor belong to $\mathcal{E}$ are added to $\mathcal{E}$ with distribution F_0.

Since the information that is used for the updates consists of independent (0,1) uniform random variables, it follows from a simple induction argument that realising the distribution of a given vertex in G_C at any time step yields a uniform (0,1) distributed random variable, independent of everything else. To be precise, this means that if we pick a vertex in G_C, its fitness necessarily has a uniform (0,1) distribution, independent of everything else, when it is fixed. Hence, all considered vertices (except the origin) are open independently and with probability p. It should now be obvious that our procedure is no different to building a site percolation cluster at the origin by the iterative method of assigning independent uniform $(0,1)$ random variables to all undetermined neighbours of the cluster and calling a vertex open if its random variable takes a value less than p. This completes the proof of Theorem 1.2. $\square$

References

[1] Athreya, K. and Ney, P., *Branching processes*, Springer, 1972

[2] Bak, P. and Sneppen, K., Punctuated equilibrium and criticality in a simple model of evolution, *Physical Review Letters*, 1993, **71**(24): 4083–4086

[3] Gillett, A., Meester, R. and van der Wal, P., Maximal avalanches in the Bak-Sneppen model, *submitted, available from* `http://www.cs.vu.nl/~rmeester/pre.html`

[4] Grimmett, G., *Percolation*, Springer, 1999, Second edition

[5] Hara, T. and Slade, G., The self-avoiding-walk and percolation critical points in high dimensions, *Combinatorics, Probability and Computing*, 1995, **4**(3): 197–215

[6] Lee, C., Zhu, X. and Gao, K., Avalanche dynamics in Bak-Sneppen evolution model observed with standard distribution width of fitness, *Nonlinearity*, 2003, **16**(1): 25–34

[7] Li, W. and Cai, X., Analytic results for scaling function and moments for a different type of avalanche in the Bak-Sneppen evolution model, *Physical Review E*, 2000, **62**(6): 7743–7747

[8] Meester, R. and Quant, C., Connections between 'classical' and 'self-organised' criticality, *to appear in Markov processes and related fields*

[9] Meester, R. and Znamenski, D., Limit behavior of the Bak-Sneppen evolution model, *Annals of Probability*, 2003, **31**(4): 1986–2002

[10] Meester, R. and Znamenski, D., Critical thresholds and the limit distribution in the Bak-Sneppen model, *Communications in Mathematical Physics*, 2004, **246**(1): 63–86