Combinatorics, Probability and Computing (2014) 23, 290–315.
doi:10.1017/S096354831300059X

Site Percolation on the *d*-Dimensional Hamming Torus

DAVID SIVAKOFF†

Department of Statistics, Department of Mathematics, Ohio State University, Columbus, OH 43210, USA
(e-mail: dsivakoff@stat.osu.edu)

Received 27 October 2011; revised 9 October 2013; first published online 29 November 2013

The d-dimensional Hamming torus is the graph whose vertices are all of the integer points inside an $a_1 n \times a_2 n \times \cdots \times a_d n$ box in $\mathbb{R}^d$ (for constants $a_1, \ldots, a_d > 0$), and whose edges connect all vertices within Hamming distance one. We study the size of the largest connected component of the subgraph generated by independently removing each vertex of the Hamming torus with probability $1 - p$. We show that if $p = \lambda/n$, then there exists $\lambda_c > 0$, which is the positive root of a degree d polynomial whose coefficients depend on $a_1, \ldots, a_d$, such that for $\lambda < \lambda_c$ the largest component has $O(\log n)$ vertices (w.h.p. as $n \to \infty$), and for $\lambda > \lambda_c$ the largest component has $(1-q)\lambda\left(\prod_i a_i\right) n^{d-1} + o(n^{d-1})$ vertices and the second largest component has $O(\log n)$ vertices w.h.p. An implicit formula for $q < 1$ is also given. The value of λ_c that we find is distinct from the critical value for the emergence of a giant component in bond percolation on the Hamming torus.

2010 *Mathematics subject classification*: Primary 05C80
Secondary 82B43

1. Introduction

Erdős and Rényi studied *bond percolation* on the complete graph in [10] (for an account of their results, see [9]). In the Erdős–Rényi model, a random subgraph of the complete graph on n vertices is obtained by independently deciding whether to remove each edge with probability $(1 - p)$ or keep it with probability p. Erdős and Rényi studied the size of the largest connected component in the random subgraph under the scaling $p = \lambda/n$, where λ is a constant parameter. They found that the size of the largest connected component is with high probability $O(\log n)$ for $\lambda < 1$, and $cn + o(n)$ for $c = c(\lambda) > 0$ when $\lambda > 1$. A sequence of events E_n is said to occur *with high probability* (w.h.p.) if $\mathbb{P}(E_n) \to 1$ as $n \to \infty$. Thus, at $\lambda = 1$, the Erdős–Rényi model undergoes a transition from having small components to having a giant connected component.

† Dissertation research at UC Davis supported in part by NSF VIGRE grant DMS-0636297, and NSF grants DMS-0805970 and DMS-0505734.

Bond percolation has since been extensively studied on other graphs $G = (V, E)$. Alternatively, random subgraphs of G can be obtained by independently removing vertices (and any edges incident to those vertices) with probability $(1 - p)$; we refer to this model as *site percolation*. Also, we will refer to the vertices in the random subgraph as *occupied*, and those that are removed we call *unoccupied*.

An example of a graph for which both bond and site percolation have been studied is the n-dimensional hypercube, Q_n, which has vertex set $V = \{0, 1\}^n$ and edge set $E = \{(x, y) \in V \times V : d(x, y) = 1\}$, where $d(x, y)$ is the Hamming distance between x and y (the number of coordinates in which they differ). Ajtai, Komlós and Szemerédi [1] proved that for bond percolation on Q_n with $p = \lambda/n$, λ constant, if $\lambda < 1$ then all connected components of the bond percolation subgraph have $O(n)$ vertices w.h.p., and if $\lambda > 1$ there is a component with $c2^n + o(2^n)$ vertices w.h.p. for $c = c(\lambda) > 0$. Bollobás, Kohayakawa and Łuczak [6] found similar behaviour for site percolation on Q_n, showing that for $p = \lambda/n$, if $\lambda < 1$ all components have at most $O(n)$ vertices w.h.p., and if $\lambda > 1$, there is a component with $c\lambda n^{-1}2^n + o(n^{-1}2^n)$ vertices for $c = c(\lambda) > 0$ w.h.p. Thus, both percolation models on Q_n have the same threshold. In fact, even the proportion of vertices in the giant component above the threshold is the same.

The intuition behind each of these results is that bond percolation on the hypercube and the complete graph, and site percolation on the hypercube look locally tree-like. Since the underlying graph is D-regular, the connected component associated with a fixed vertex can be compared to a branching process with a Binomial(D, p) offspring distribution. A giant connected component exists precisely when the branching process has positive survival probability, which is when the expected number of offspring per individual ($Dp = \lambda$) exceeds 1. Borgs, Chayes, van der Hofstad, Slade and Spencer [7, 8] showed that, for a wide variety of finite transitive graphs, the threshold for the existence of a giant component in bond percolation should be at $Dp = 1$, though they focus on behaviour in the critical window, and do not give a lower bound for the size of the supercritical giant component. The case of site percolation is less clear in general, due to local dependences between edges.

We study site percolation on the *Hamming torus*, $\mathcal{H} = \mathcal{H}(n, a_1, \cdots, a_d)$, which is the graph given by the vertex and edge sets

$$\begin{aligned} V =& \{(x_1, \ldots, x_d) \in \mathbb{Z}^d : 1 \leqslant x_i \leqslant a_i n,\ i = 1, \ldots, d\}, \\ E =& \{(x, y) \in V \times V : d(x, y) = 1\}, \end{aligned}$$

where $d(x, y)$ is the Hamming distance between x and y. That is, two vertices in the $a_1 n \times a_2 n \times \cdots \times a_d n$ box in $\mathbb{Z}^d$ are adjacent if and only if they differ in exactly one coordinate. For $d = 1$ this is the complete graph on $a_1 n$ vertices, and for $d = 2$ this is the rook graph on an $a_1 n$ by $a_2 n$ chessboard. Site percolation on $\mathcal{H}$ is the random subgraph $\mathcal{H}_p$, obtained by independently deciding whether to discard each vertex with probability $(1 - p)$.

This model is amenable to comparison with a multitype branching process, $\mathcal{X}_{a_1,\ldots,a_d,\lambda,n} = (\boldsymbol{Z}_t)_{t=1}^{\infty}$, with d types, so $\boldsymbol{Z}_t = (Z_t^1, \ldots, Z_t^d)$. In the first generation of this process there are $Z_1^j \sim$ Binomial$(a_j n, \lambda/n)$ offspring of type j for each j. Then, letting $\mathbf{e}_i$ denote the ith

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

standard basis vector in $\mathbb{R}^d$, in each subsequent generation an individual of type i has $(Z_{t+1}^j \mid \boldsymbol{Z}_t = \mathbf{e}_i) \sim \text{Binomial}(a_j n, \lambda/n)$ offspring of type j for each $j \neq i$ and zero offspring of type i. This branching process has non-zero survival probability if and only if the largest eigenvalue of the matrix of expected progeny, $\boldsymbol{M}_\lambda$, is strictly larger than one [2]. Here,

$$\boldsymbol{M}_\lambda = \begin{pmatrix} 0 & \lambda a_2 & \lambda a_3 & \cdots & \lambda a_d \\ \lambda a_1 & 0 & \lambda a_3 & \cdots & \lambda a_d \\ \lambda a_1 & \lambda a_2 & 0 & & \vdots \\ \vdots & \vdots & & \ddots & \lambda a_d \\ \lambda a_1 & \lambda a_2 & \cdots & \lambda a_{d-1} & 0 \end{pmatrix}, \tag{1.1}$$

where the entry in position (i, j) is the expected number of type j offspring to which a type i individual will give birth. This branching process is a good local approximation for a connected component of $\mathcal{H}_p$ that contains a designated vertex when $p = \lambda/n$, which motivates the following theorems.

Theorem 1.1. *Fix $d \geqslant 2$ and $a_1, \ldots, a_d > 0$, and let λ_c be the unique positive solution to $\det(\boldsymbol{I} - \boldsymbol{M}_{\lambda_c}) = 0$. If $\lambda < \lambda_c$ is constant and $p = \lambda/n$, then the size of the largest connected component of $\mathcal{H}_p$ is $O(\log n)$ w.h.p.*

Theorem 1.2. *Fix $d \geqslant 2$ and $a_1, \ldots, a_d > 0$, and let λ_c be the unique positive solution to $\det(\boldsymbol{I} - \boldsymbol{M}_{\lambda_c}) = 0$. If $\lambda > \lambda_c$ is constant and $p = \lambda/n$, then there is a constant $q = q(\lambda)$ such that the size of the largest connected component of $\mathcal{H}_p$ is*

$$(1-q)\lambda\left(\prod_i a_i\right) n^{d-1}(1 + o(1)) \quad \text{w.h.p.}$$

Furthermore, the second largest component has $O(\log n)$ vertices w.h.p.

The constant $q = q(\lambda)$ appearing in Theorem 1.2 is the extinction probability of the branching process corresponding to $\mathcal{X}_{a_1,\ldots,a_d,\lambda,n}$ in the limit as $n \to \infty$, so the number of offspring of each type follows an independent Poisson distribution, and q satisfies $0 < q < 1$ for $\lambda > \lambda_c$. It is defined precisely at the start of Section 3. In both theorems $\boldsymbol{I}$ is the $d \times d$ identity matrix. Note that the total number of vertices in site percolation on the Hamming torus is asymptotically the expected number, $p|V| = \lambda\left(\prod_i a_i\right) n^{d-1}$, so $(1-q)$ is the proportion of vertices remaining that are in the giant component. The critical value, λ_c, can be defined equivalently as the root of a degree d polynomial, which is specified by the following lemma.

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

Lemma 1.3. *Fix $d \geqslant 2$ and $a_1, \dots, a_d > 0$, and define $\boldsymbol{M}_\lambda$ as in* (1.1). *Then the following holds:*

$$\det(\boldsymbol{I} - \boldsymbol{M}_\lambda) = 1 - \sum_{\ell=2}^{d} (\ell - 1)\lambda^\ell \sum_{\substack{S \subset \{1,\dots,d\} \\ |S|=\ell}} \prod_{i \in S} a_i.$$

Proof. We evaluate the determinant as

$$\det(\boldsymbol{I} - \boldsymbol{M}_\lambda) = \sum_{\sigma \in \mathfrak{S}_d} \operatorname{sign}(\sigma) \prod_{i=1}^{d} (\boldsymbol{I} - \boldsymbol{M}_\lambda)_{i\sigma(i)},$$

where $\mathfrak{S}_d$ is the set of permutations of $\{1, \dots, d\}$. We break this sum up according to the diagonal elements selected in the product (fixed points of σ). For $S \subset \{1, \dots, d\}$ with $|S| = \ell$, the sum of the terms such that the diagonal elements selected are precisely those indexed by S^c is equal to the determinant of the submatrix of $-\boldsymbol{M}_\lambda$ obtained by selecting the rows and columns indexed by S. Taking a factor from each column, this is equal to $(-\lambda)^\ell \prod_{i \in S} a_i$ times the determinant of the ℓ-by-ℓ matrix with zeros on the diagonal and ones off the diagonal, which is easily seen to be $(-1)^{\ell-1}(\ell - 1)$. Summing over S gives the result. □

Note that $\det(\boldsymbol{I} - \boldsymbol{M}_\lambda)$ is strictly decreasing in λ for $\lambda > 0$, is positive when $\lambda = 0$ and tends to $-\infty$ as $\lambda \to \infty$. Thus, the equation $\det(\boldsymbol{I} - \boldsymbol{M}_\lambda) = 0$ has a unique positive solution λ_c. Since $\det(\boldsymbol{I} - \boldsymbol{M}_\lambda) = 0$ if and only if 1 is an eigenvalue of $\boldsymbol{M}_\lambda$, and the eigenvalues of $\boldsymbol{M}_\lambda$ scale with λ, it follows that the largest eigenvalue of $\boldsymbol{M}_\lambda$ is less than 1 for $\lambda < \lambda_c$ and greater than 1 for $\lambda > \lambda_c$.

Random *edge* subgraphs of the Hamming torus were considered by Borgs, Chayes, van der Hofstad, Slade and Spencer [7, 8], who indicated that the threshold, scaled by n, for the emergence of a giant component in bond percolation should be $1/(a_1 + \cdots + a_d)$, though they do not provide a lower bound for the size of the largest component above the threshold. This lower bound was later proved by van der Hofstad and Luczak [11] in the case $d = 2$ and $a_1 = a_2 = 1$, demonstrating that the threshold occurs at $1/2$. When $d = 2$, the site percolation threshold is $\lambda_c = 1/\sqrt{a_1 a_2}$, so the two processes clearly differ.

A model similar to site percolation on the Hamming torus, called line-of-sight percolation, was studied by Bollobás, Janson and Riordan [5]. They studied site percolation on the graph with vertex set $\mathbb{Z}^2$ in which two vertices are neighbours if they agree in one coordinate and differ by at most $\omega \geqslant 1$ in the other. They showed that if $p_c(\omega)$ is the critical probability for site percolation on this graph, then $\lim_{\omega \to \infty} \omega p_c(\omega) = \log(3/2)$. They also showed that the analogous result holds for the emergence of a giant component on the $n \times n$ grid when $\omega = o(n)$, and they discuss the critical threshold when $\omega = cn$.

Multitype branching processes have also been employed in the analysis of *inhomogeneous* bond percolation of the complete graph by Bollobás, Janson and Riordan [4]. In the inhomogeneous model, the probability of retaining the edge between vertices i and j is p_{ij}, and the p_{ij} may not be equal. However, the inclusion or exclusion of each edge still occurs independently of all other edges. Since the number of neighbours that a vertex

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

has depends on the probabilities of the edges incident to that vertex, the inhomogeneous model looks locally like a multitype branching process: each vertex has a different offspring distribution, and is thus of a different 'type'. Bollobás, Janson and Riordan [4] proved that, under certain conditions on the values of the p_{ij}, a giant component exists with high probability precisely when the corresponding multitype branching process has positive probability of survival. For site percolation on the Hamming torus, the connection to a multitype branching process arises in a different manner. As in [5], when exploring the neighbourhood of a vertex, each direction must be searched except for the direction from which a vertex was discovered. Thus, each new vertex is given a type according to the direction from which it was discovered.

In Section 2 we prove Theorem 1.1 by coupling a process revealing the vertices in a connected component of $\mathcal{H}_p$ with the multitype branching process $\mathcal{X}_{a_1,\ldots,a_d,\lambda,n}$. In Section 3 we prove Theorem 1.2 in four steps. The first step (Section 3.1) is to show that the process of revealing the vertices in a connected component of $\mathcal{H}_p$ will either terminate before discovering $O(\log n)$ occupied vertices, or will reveal $m = \Theta(\log n)$ occupied vertices with unrevealed neighbourhoods with high probability. The second step (Section 3.2) is to show that if the process of revealing vertices acquires m occupied vertices with unrevealed neighbourhoods, then it can be coupled with a lower-bounding branching process, which will reach size $n^{d-4/3}$ with high probability. The third step (Section 3.3) is to show that any two component discovering processes (started from two different vertices in V) that reach size $n^{d-4/3}$ will join together with high probability. For this step, we use a slight modification of a 'sprinkling' argument like the one used in [1], but which gives us better control over the event that a vertex is in the giant component. The final step (Section 3.4) is to show that the proportion of vertices in components of size $O(\log n)$ converges to q in probability.

2. Subcritical behaviour

Theorem 1.1 comes from a direct comparison with the binomial multitype branching process described above, and thus follows from Proposition 2.2 below, which bounds the total size of a multitype branching process. Consider a time-homogeneous multitype branching process $\mathcal{X} = (\boldsymbol{Z}_t)_{t=0}^{\infty}$ with d types, so $\boldsymbol{Z}_t = (Z_t^1, Z_t^2, \ldots, Z_t^d)$, where Z_t^i is the number of type i individuals in the tth generation. $\boldsymbol{Z}_0$ is assumed to be deterministic, and all birth events are independent of one another. Let $\boldsymbol{M} = (m_{ij})$ denote the matrix of expectations,

$$m_{ij} = \mathbb{E}\big(Z_1^j \mid \boldsymbol{Z}_0 = \mathbf{e}_i\big). \tag{2.1}$$

We say that a matrix, $\boldsymbol{A}$, or a vector, $\boldsymbol{x}$, is *positive* if $A_{ij} > 0$ for all i and j or $x_i > 0$ for all i. Lemma 2.1 is a simplified restatement of the Perron–Frobenius theorem as it appears in [2].

Lemma 2.1. *If $\boldsymbol{M}^N$ is positive for some natural number N, then $\boldsymbol{M}$ has a positive simple eigenvalue, ρ, that is greater in absolute value than any other eigenvalue, and ρ corresponds to a positive right eigenvector $\boldsymbol{\mu} : \boldsymbol{M}\boldsymbol{\mu} = \rho\boldsymbol{\mu}$.* □

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

Proposition 2.2 bounds the total size of a subcritical branching process with exponentially decaying tails, and is surely known in some form (a proof sketch for a similar and more general result appears in the proof of Theorem 12.5 of [4]). We were unable to find a full proof of the specific form that we use, so we include a short proof here. When $d \geqslant 3$, it is easy to check that $\boldsymbol{M}_\lambda^2$ is positive, so Lemma 2.1 applies, and when $d = 2$ it can be verified directly that $\boldsymbol{M}_\lambda$ has a unique positive eigenvalue corresponding to a positive right eigenvector. Thus, $\boldsymbol{M}_\lambda$ will satisfy the condition on $\boldsymbol{M}$ in Proposition 2.2.

Proposition 2.2. *Let $\boldsymbol{M}$ be the expectation matrix of a multitype branching process $\mathcal{X}$ as in* (2.1). *Suppose there exists $0 < \rho < 1$ and $\boldsymbol{\mu}$ positive such that $\boldsymbol{M}\boldsymbol{\mu} = \rho\boldsymbol{\mu}$, and there exists $\theta_0 > 0$ such that $\mathbb{E}\left[e^{\theta_0\langle \boldsymbol{Z}_1, \boldsymbol{\mu}\rangle} \mid \boldsymbol{Z}_0 = \mathbf{e}_i\right] < \infty$ for $i = 1, \ldots, d$. Then*

$$\mathbb{P}\left(\sum_{t=0}^{\infty} \|\boldsymbol{Z}_t\|_1 > x\right) \leqslant Ce^{-\alpha x},$$

where $\alpha, C > 0$ can be chosen as

$$e^{-\alpha} := \min_{\theta \in [0,\theta_0]} \max_i \ \mathbb{E}\left[e^{\theta\langle \boldsymbol{Z}_1 - \mathbf{e}_i, \boldsymbol{\mu}\rangle} \mid \boldsymbol{Z}_0 = \mathbf{e}_i\right],$$

and, letting θ' be the smallest value of θ for which this minimum is attained, $C := e^{\theta'\langle \boldsymbol{Z}_0, \mu\rangle}$. □

Proof. Consider a random walk version of the multitype branching process, $\boldsymbol{S}_t = (S_t^1, \ldots, S_t^d)$, constructed as follows. $\boldsymbol{S}_0 = \boldsymbol{Z}_0$ is a non-random initial vector whose ith component indicates the number of type i individuals who are *active*. At each step, an active individual is chosen (depending only on the current state $\boldsymbol{S}_t$), it gives birth to a random number of individuals depending on its type and according to the law for that type in the branching process (these new individuals are considered active), and then it is made inactive (thus no longer included in $\boldsymbol{S}_t$). If J_t is the random variable that takes the value i if an active individual of type i is selected at time t, then

$$\boldsymbol{S}_{t+1} = \boldsymbol{S}_t + \sum_{i=1}^{d} \mathbb{1}_{\{J_{t+1}=i\}}\left(\boldsymbol{X}_{t+1}^i - \mathbf{e}_i\right), \tag{2.2}$$

where for each $i = 1, \ldots, d$ and $t \geqslant 0$ the random vectors $\boldsymbol{X}_t^i$ are independent and equal in distribution to $(\boldsymbol{Z}_1 \mid \boldsymbol{Z}_0 = \mathbf{e}_i)$. Notice that J_{t+1} is dependent on $\boldsymbol{S}_t$, but $\boldsymbol{X}_{t+1}^i$ is not. The process continues until the stopping time $T := \inf\{t : \boldsymbol{S}_t = \mathbf{0}\}$, at which time the process dies out, and we have

$$T \overset{d}{=} \sum_{t=0}^{\infty} \|\boldsymbol{Z}_t\|_1.$$

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

Consider an increment of this random walk, and let

$$\begin{aligned}\phi_{\boldsymbol{v}}(\theta) &:= \mathbb{E}\big[\exp\big(\theta\langle \boldsymbol{S}_{t+1}-\boldsymbol{S}_t,\boldsymbol{\mu}\rangle\big) \mid \boldsymbol{S}_t=\boldsymbol{v}\big]\\ &= \sum_i \mathbb{P}(J_{t+1}=i \mid \boldsymbol{S}_t=\boldsymbol{v})\ \mathbb{E}e^{\theta\langle (\boldsymbol{X}^i_{t+1}-\boldsymbol{e}_i),\boldsymbol{\mu}\rangle}\\ &\leqslant \max_i\ \mathbb{E}e^{\theta\langle (\boldsymbol{X}^i_1-\boldsymbol{e}_i),\boldsymbol{\mu}\rangle}\\ &=: \psi(\theta). \end{aligned} \tag{2.3}$$

Note that the last expression does not depend on the previous state vector, $\boldsymbol{v}$. By our assumption, there exists $\theta_0>0$ so that $\psi(\theta_0)<\infty$, so we have

$$\frac{d}{d\theta}\big(\mathbb{E}e^{\theta\langle (\boldsymbol{X}^i_1-\boldsymbol{e}_i),\boldsymbol{\mu}\rangle}\big)|_{\theta=0} = \mu_i(\rho-1).$$

Since we assumed $\rho<1$ and $\mu_i>0$ for all i, we have that $\psi(\theta)<1$ for θ near zero. Further, since ψ is continuous on $[0,\theta_0]$, there exist $\theta',\alpha>0$ such that $\min_{\theta\in[0,\theta_0]}\psi(\theta)=\psi(\theta')=e^{-\alpha}<1$. It is easy to see that $e^{\theta\langle \boldsymbol{S}_t,\mu\rangle}/\psi(\theta)^t$ is a positive supermartingale by (2.3), so, by the Optional Stopping Theorem,

$$C := e^{\theta'\langle \boldsymbol{S}_0,\boldsymbol{\mu}\rangle} \geqslant \mathbb{E}\left[\frac{e^{\theta'\langle \boldsymbol{S}_T,\boldsymbol{\mu}\rangle}}{\psi(\theta')^T}\right] = \mathbb{E}e^{\alpha T},$$

and then by Markov's inequality,

$$\mathbb{P}(T>x)=\mathbb{P}(e^{\alpha T}>e^{\alpha x}) \leqslant Ce^{-\alpha x}. \qquad \square$$

We now prove the following lemma, which immediately implies Theorem 1.1. We will make use of this slightly stronger statement in Section 3.2 of the proof of Theorem 1.2.

Lemma 2.3. *Fix $d\geqslant 2$ and $a_1,\dots,a_d>0$, let λ_c be the unique positive solution to $\det(\boldsymbol{I}-\boldsymbol{M}_{\lambda_c})=0$, and let $\lambda<\lambda_c$ be constant. If $p=\lambda/n$, then for any $\eta>0$ there exists $\beta>0$ such that, for all sufficiently large n, with probability at least $1-n^{-\eta}$, the size of the largest connected component of $\mathcal{H}_p$ is at most $\beta\log n$.*

Proof. Consider a fixed vertex, $\boldsymbol{v}$, in the Hamming torus, and let $C_{\boldsymbol{v}}$ be the cluster containing that vertex ($C_{\boldsymbol{v}}=\emptyset$ if the vertex $\boldsymbol{v}$ is unoccupied). We can reveal the vertices in $C_{\boldsymbol{v}}$ using the Cluster Discovering Algorithm that follows. As in [9], we will use three sets to keep track of the status of each vertex in the process: U_t denotes the set of *unseen* vertices, A_t is the set of *active* vertices (occupied vertices whose neighbourhoods have not yet been searched entirely), and R_t denotes the set of *removed* vertices (occupied vertices whose neighbourhoods have been exhausted). In the language of epidemics, this is like an SIR epidemic, where U_t is the susceptible set, A_t is the infected set and R_t is the recovered set. Our process differs from an epidemic in that $U_t\cup A_t\cup R_t\subsetneq V$ for $t\geqslant 1$, since not every vertex that is seen (and thus leaves U_t) becomes active. In particular, those vertices that are observed but not occupied are unaccounted for.

To define the algorithm, we let $\mathcal{N}(\boldsymbol{v}) := \{\boldsymbol{w}\in V : d(\boldsymbol{v},\boldsymbol{w})=1\}$ denote the *neighbourhood* of the vertex $\boldsymbol{v}\in V$.

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

Cluster Discovering Algorithm.

(1) Initialize the sets $R_0 = \emptyset$, $A_0 = \{\boldsymbol{v}\}$, $U_0 = V \setminus A_0$.

(2) Choose any vertex, $\boldsymbol{v}_t$, from A_t.

(3) $A_{t+1} = A_t \setminus \{\boldsymbol{v}_t\} \cup \{\boldsymbol{w} \in \mathcal{N}(\boldsymbol{v}_t) \cap U_t : \boldsymbol{w} \text{ is occupied}\}$.

(4) $U_{t+1} = U_t \setminus \mathcal{N}(\boldsymbol{v}_t)$.

(5) $R_{t+1} = R_t \cup \{\boldsymbol{v}_t\}$.

(6) If A_{t+1} is empty, then return $C_{\boldsymbol{v}} = R_{t+1}$, otherwise increment t and go to step (2).

In words, at time t, we select a vertex, $\boldsymbol{v}_t$, from the set of active vertices, A_t, and search its unseen neighbourhood, $\mathcal{N}(\boldsymbol{v}_t) \cap U_t$. We add each occupied vertex in the unseen neighbourhood of $\boldsymbol{v}_t$ to A_{t+1}, remove *all* of the neighbours of $\boldsymbol{v}_t$ from U_{t+1} (since they have now been inspected), and move $\boldsymbol{v}_t$ from A_{t+1} to R_{t+1}. Thus, at the end of the process, when A_{t+1} is empty, the set R_{t+1} is equal to $C_{\boldsymbol{v}}$.

To compare this process with a multitype branching process, we say that at step (3), a neighbour of $\boldsymbol{v}_t$, say $\boldsymbol{w}$, is *of type* i if $\boldsymbol{v}_t - \boldsymbol{w} = m\boldsymbol{e}_i$ for some integer m. If $\boldsymbol{v}_t$ ($t \geqslant 1$) is of type i, then only vertices of types $j \in \{1, \dots, d\} \setminus \{i\}$ may become active, since all of $\boldsymbol{v}_t$'s neighbours in the $\mathbf{e}_i$ direction will have been removed from U_{t+1} when $\boldsymbol{v}_t$ became active. In the first iteration, however, $\boldsymbol{v}$ has unseen neighbours in each of the d directions. Since $\boldsymbol{v}_t$ has at most $a_j n$ unseen neighbours in the $\mathbf{e}_j$ direction, $|C_{\boldsymbol{v}}|$ is stochastically dominated by the total size of the multitype branching process $\mathcal{X}_{a_1,\dots,a_d,\lambda,n}$. In turn, $\mathcal{X}_{a_1,\dots,a_d,\lambda,n}$ is stochastically dominated by the time-homogeneous branching process $\mathcal{X}'_{a_1,\dots,a_d,\lambda,n}$ with the same d types as $\mathcal{X}_{a_1,\dots,a_d,\lambda,n}$, but with the initial state $\boldsymbol{Z}_0 := (1, \dots, 1)$, so the first generation has Binomial$((d-1)a_i n, \lambda/n)$ type i individuals. The expectation matrix for this branching process is $\boldsymbol{M}_\lambda$, and the largest eigenvalue of $\boldsymbol{M}_\lambda$ is less than 1 if $\lambda < \lambda_c$. In this regime, we have that the moment generating functions for $(\langle \boldsymbol{Z}_1 - \mathbf{e}_i, \boldsymbol{\mu} \rangle \mid \boldsymbol{Z}_0 = \mathbf{e}_i)$ are uniformly (in n) bounded above for $\theta \geqslant 0$, as

$$\begin{aligned}
\mathbb{E}\left[e^{\theta\langle \boldsymbol{Z}_1 - \mathbf{e}_i, \boldsymbol{\mu}\rangle} \mid \boldsymbol{Z}_0 = \mathbf{e}_i\right] &= e^{-\theta\mu_i} \prod_{j \neq i} \left[1 + \frac{\lambda}{n}(e^{\theta\mu_j} - 1)\right]^{a_j n} \\
&\leqslant e^{-\theta\mu_i} \prod_{j \neq i} \exp\left[\lambda a_j (e^{\theta\mu_j} - 1)\right] \\
&=: \psi_i(\theta). \qquad (2.4)
\end{aligned}$$

It is easy to verify that $\psi_i'(0) = \mu_i(\rho - 1) < 0$, $\psi_i(0) = 1$, $\psi_i(\theta) \to \infty$ as $\theta \to \infty$, and $\psi_i''(\theta) > 0$ for all $\theta > 0$, which imply that we can choose $\alpha, C > 0$ such that

$$\begin{aligned}
e^{-\alpha} &:= \min_{\theta \geqslant 0} \max_i \ \psi_i(\theta) =: \max_i \ \psi_i(\hat{\theta}), \\
C &:= \exp\left[\hat{\theta}\|\boldsymbol{\mu}\|_1\right].
\end{aligned}$$

The inequality in (2.4) shows that the corresponding generating functions for $\boldsymbol{Z}_t$ are bounded uniformly in n by ψ_i. Applying Proposition 2.2 to $\mathcal{X}'_{a_1,\dots,a_d,\lambda,n}$ with $\theta_0 = \hat{\theta}$, we

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

obtain C_n, α_n such that, by (2.4), $e^{-\alpha_n} \leqslant e^{-\alpha}$ and $C_n \leqslant C$ for all n, and

$$\mathbb{P}\left(|C_{\boldsymbol{v}}| > \frac{d+\eta+1}{\alpha}\log n\right) \leqslant \mathbb{P}\left(\sum_{t=0}^{\infty}\|\boldsymbol{Z}_t\|_1 > \frac{d+\eta+1}{\alpha}\log n\right) \leqslant Cn^{-d-\eta-1}.$$

This implies the desired result with $\beta = (d+\eta+1)/\alpha$:

$$\mathbb{P}\left(\max_{\boldsymbol{v}\in V}|C_{\boldsymbol{v}}| > \frac{d+\eta+1}{\alpha}\log n\right) \leqslant C\left(\prod_{i=1}^{d}a_i\right)n^{-\eta-1} = o(n^{-\eta}). \qquad \square$$

3. Supercritical behaviour

Theorem 1.2 shows that λ_c is the critical threshold for the emergence of a giant component. If we let $q_i = \mathbb{P}\big(\boldsymbol{Z}_t = \boldsymbol{0} \text{ for some } t \mid \boldsymbol{Z}_0 = \mathbf{e}_i\big)$ be the extinction probabilities for the multitype branching process in which an individual of type i gives birth to a Poisson(λa_j) number of type j individuals for $j \neq i$ and zero individuals of type i, then the proportion of occupied vertices in the largest component when $\lambda > \lambda_c$ is

$$1 - \left(\prod_i q_i\right)^{1/(d-1)} \equiv 1 - q > 0.$$

Theorem 3.1 (Theorem 1.2 restated). *Fix $d \geqslant 2$ and $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_d > 0$, and let λ_c be the unique positive solution to $\det(\boldsymbol{I} - \boldsymbol{M}_{\lambda_c}) = 0$. If $\lambda > \lambda_c$ and $p = \lambda/n$, then the size of the largest connected component of site percolation on the Hamming torus is*

$$(1-q)\lambda\left(\prod_i a_i\right)n^{d-1} + o(n^{d-1}) \quad \textit{w.h.p.}$$

Furthermore, the second largest component has $O(\log n)$ vertices w.h.p.

The proof of Theorem 1.2 begins here, and spans four subsections, Sections 3.1–3.4. The proof is summarized at the end of Section 3.4.

We begin by considering $C_{\boldsymbol{v}}$, the cluster containing a fixed vertex $\boldsymbol{v}$, and reveal the vertices in $C_{\boldsymbol{v}}$ using a similar process to the Cluster Discovering Algorithm, but with one noteworthy modification. We will remove some additional vertices from the unseen set, U_{t+1}, at step (4) before they can be observed in any subsequent iterations. In particular, we remove vertices that are neighbours of two or more vertices in the active set, A_{t+1}. This will avoid the problem of generating closed loops, which would severely reduce the unseen neighbourhoods of active vertices and make it difficult to couple the process with a lower-bounding branching process. In [5], a similar idea was used where each newly discovered vertex 'claims' its neighbours. Here, if two active vertices could discover the same unseen vertex, we remove the unseen vertex completely, so neither active vertex claims this neighbour. The result is that this algorithm only returns a subset of $C_{\boldsymbol{v}}$, but it will suffice because we will be able to bound the number of vertices that are lost in this manner.

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

Cluster Discovering Algorithm (modified).

(1) Initialize the sets $R_0 = \emptyset$, $A_0 = \{\boldsymbol{v}\}$, $U_0 = V \setminus A_0$.
(2) Choose any vertex, $\boldsymbol{v}_t$, from A_t.
(3) $A_{t+1} = A_t \setminus \{\boldsymbol{v}_t\} \cup \{\boldsymbol{w} \in \mathcal{N}(\boldsymbol{v}_t) \cap U_t : \boldsymbol{w} \text{ is occupied}\}$.
(4) $U_{t+1} = \big(U_t \setminus \mathcal{N}(\boldsymbol{v}_t)\big) \setminus \{\boldsymbol{w} \in \mathcal{N}(\boldsymbol{u}) \cap \mathcal{N}(\boldsymbol{v}) : \boldsymbol{u}, \boldsymbol{v} \in A_{t+1}, \boldsymbol{u} \neq \boldsymbol{v}\}$.
(5) $R_{t+1} = R_t \cup \{\boldsymbol{v}_t\}$.
(6) If A_{t+1} is empty, then return R_{t+1}, otherwise increment t and go to step (2).

During this process, we will need to keep track of the number of active vertices of each type, so for $t \geqslant 1$ let $\mathcal{A}_t^{(i)}$ denote the number of type i vertices in A_t, let $\boldsymbol{\mathcal{A}}_t = \big(\mathcal{A}_t^{(1)}, \ldots, \mathcal{A}_t^{(d)}\big)$, and let $\mathcal{N}^i(\boldsymbol{v}) = \{\boldsymbol{w} \in V \mid \boldsymbol{v} - \boldsymbol{w} = m\mathbf{e}_i \text{ for some } m \in \mathbb{Z}\}$. As in the proof of Theorem 1.1, a vertex, $\boldsymbol{w}$, discovered at time t is labelled type i if $\boldsymbol{w} \in \mathcal{N}^i(\boldsymbol{v}_t)$. That is, a vertex is of type i if it is discovered by searching neighbouring vertices in the direction of the ith basis vector. Note that we do not define $\boldsymbol{\mathcal{A}}_0$ in the same way because the initial vector, $\boldsymbol{v}_0$, is not of any of the d types, since it has unseen neighbours in all d directions. For convenience of later calculations, we will assume $\|\boldsymbol{\mathcal{A}}_0\|_1 := 1$. This will not hinder us in any significant way, and really only comes into play in the computation of q later on.

The above algorithm will yield a lower bound on $C_{\boldsymbol{v}}$, but we will make use of an upper-bounding branching process, as in the proof of Theorem 1.1, to tightly control the size of $C_{\boldsymbol{v}}$. We can couple the random walk version of the upper-bounding branching process with $\boldsymbol{\mathcal{A}}_t$ by using the same random variables in the construction of $\boldsymbol{S}_t$ as in the construction of $\boldsymbol{\mathcal{A}}_t$ whenever possible, and when independence is an issue, we add independent copies of these random variables to $\boldsymbol{S}_t$. More rigorously, let $\xi_{\boldsymbol{v}}$ be the random variable that takes the value 1 if the vertex $\boldsymbol{v}$ is occupied and is 0 otherwise. Define $\|\boldsymbol{S}_0\|_1 = 1$, and for each $i \in \{1, \ldots, d\}$

$$S_1^{(i)} = \sum_{\boldsymbol{w} \in \mathcal{N}^i(\boldsymbol{v}_0) \cap U_0} \xi_{\boldsymbol{w}}.$$

We define $\boldsymbol{S}_t = (S_t^{(1)}, \ldots, S_t^{(d)})$ iteratively for each $t \geqslant 1$ and $i \in \{1, \ldots, d\}$ as

$$S_{t+1}^{(i)} = S_t^{(i)} + \mathbb{1}_{\{\boldsymbol{v}_t \text{ is not of type } i\}} \left[\sum_{\boldsymbol{w} \in \mathcal{N}^i(\boldsymbol{v}_t) \cap U_t} \xi_{\boldsymbol{w}} + \sum_{k=1}^{a_i n - |\mathcal{N}^i(\boldsymbol{v}_t) \cap U_t|} \eta_k^{(t,i)} \right] - \mathbb{1}_{\{\boldsymbol{v}_t \text{ is of type } i\}}, \tag{3.1}$$

where the $\eta_k^{(t,i)}$ are i.i.d. Bernoulli(p) random variables, and are independent of the $\xi_{\boldsymbol{w}}$. The additional random variables, $\eta_k^{(t,i)}$, are added in so that $\boldsymbol{S}_t$ will be a random walk with i.i.d. increments. As long as $\boldsymbol{\mathcal{A}}_t \neq \mathbf{0}$, then $\mathcal{A}_t^{(i)} \leqslant S_t^{(i)}$ for each i, so the same type of individual can be chosen for both processes. If $\boldsymbol{\mathcal{A}}_t = \mathbf{0}$ and $\boldsymbol{S}_t \neq \mathbf{0}$, then we can choose a type such that $S_t^{(i)} > 0$ for the next increment of the process $\boldsymbol{S}_t$, and let this determine the value of the indicators that $\boldsymbol{v}_t$ is of type i or not. From this construction, it is clear that if a type i individual is chosen at time $t \geqslant 1$, then

$$\boldsymbol{S}_{t+1} - \boldsymbol{S}_t \stackrel{d}{=} (\boldsymbol{Z}_2 \mid \boldsymbol{Z}_1 = \mathbf{e}_i) - \mathbf{e}_i,$$

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

where $\boldsymbol{Z}_t$ is the branching process $\mathcal{X}_{a_1,\dots,a_d,\lambda,n}$. Therefore we have

$$\inf\{t : \boldsymbol{\mathcal{A}}_t = \mathbf{0}\} \leqslant \inf\{t : \boldsymbol{S}_t = \mathbf{0}\} \stackrel{d}{=} \sum_t \|\boldsymbol{Z}_t\|_1.$$

3.1. Establishing a set of $m = b \log n$ active vertices

As in the proof of Theorem 1.1, we make use of an upper-bounding branching process, but this time in a more intimate fashion. We couple the random walk version of the upper-bounding branching process, $\boldsymbol{S}_t$, with $\boldsymbol{\mathcal{A}}_t$ as in equation (3.1). Recall from equation (2.2) that we can also write

$$\boldsymbol{S}_{t+1} \stackrel{d}{=} \boldsymbol{S}_t + \sum_{i=1}^{d} \mathbb{1}_{\{J_{t+1}=i\}} \big(\boldsymbol{X}^i_{t+1} - \mathbf{e}_i\big),$$

where J_{t+1} is the random variable that takes the value i if $\boldsymbol{v}_t$ is of type i (or, when $\boldsymbol{\mathcal{A}}_t = \mathbf{0}$, if a type i individual is chosen at time $t+1$ in the random walk process), and where the $\boldsymbol{X}^i_t$ are independent and distributed as $\boldsymbol{Z}_1$ conditional on $\boldsymbol{Z}_0 = \boldsymbol{e}_i$. For this process (and later for a lower-bounding process, $\boldsymbol{W}_t$) we will need the following large deviation bounds.

Lemma 3.2. *Let $\boldsymbol{S}_t$ be the random walk version of a branching process with d types in which an individual of type i has a* Binomial$(a_j n, \lambda/n)$ *number of offspring of type $j \neq i$ and zero offspring of type i. Suppose $\boldsymbol{M}_\lambda \boldsymbol{\mu} = \rho \boldsymbol{\mu}$, where $\rho > 1$ and $\boldsymbol{\mu}$ is the corresponding positive eigenvector normalized so that $\|\boldsymbol{\mu}\|_1 = 1$, and $\mu_{\max} := \max_i(\mu_i)$, $\mu_{\min} := \min_i(\mu_i)$. Then, given $y < 1 < x$, there exist $\eta_1, \eta_2 > 0$ such that*

$$\mathbb{P}\big(\langle \boldsymbol{S}_t - \boldsymbol{S}_0, \boldsymbol{\mu}\rangle \geqslant x(\rho-1)\mu_{\max}\, t\big) \leqslant e^{-\eta_1 t}, \tag{3.2}$$

$$\mathbb{P}\big(\langle \boldsymbol{S}_t - \boldsymbol{S}_0, \boldsymbol{\mu}\rangle \leqslant y(\rho-1)\mu_{\min} t,\ \|\boldsymbol{S}_t\| > 0\big) \leqslant e^{-\eta_2 t}. \tag{3.3}$$

Proof. We will first prove inequality (3.2). By equation (2.3), the moment generating function for an increment of the process $\langle \boldsymbol{S}_t - \boldsymbol{S}_0, \boldsymbol{\mu}\rangle$ is bounded, for $\boldsymbol{v} > \mathbf{0}$, by

$$\begin{aligned}\phi_{\boldsymbol{v}}(\theta) &:= \mathbb{E}\big[\exp\big(\theta\langle \boldsymbol{S}_{t+1} - \boldsymbol{S}_t, \boldsymbol{\mu}\rangle\big) \mid \boldsymbol{S}_t = \boldsymbol{v}\big] \\ &\leqslant \max_i\ \mathbb{E}e^{\theta\langle (\boldsymbol{X}^i_1 - \boldsymbol{e}_i), \boldsymbol{\mu}\rangle} =: \psi(\theta).\end{aligned}$$

Since $\mathbf{0}$ is an absorbing state for $\boldsymbol{S}_t$, all future increments will be $\mathbf{0}$, and we have $\phi_{\mathbf{0}}(\theta) \equiv 1$. To avoid this case, we only consider the event that the process has not yet died by time t. Letting $T := \inf\{t : \boldsymbol{S}_t = \mathbf{0}\}$, and using the above estimate, we see that $e^{\theta\langle \boldsymbol{S}_t, \boldsymbol{\mu}\rangle} \mathbb{1}_{\{T > t-1\}} [\psi(\theta)]^{-t}$ is a supermartingale, which implies that

$$\mathbb{E}e^{\theta\langle \boldsymbol{S}_t - \boldsymbol{S}_0, \boldsymbol{\mu}\rangle} \mathbb{1}_{\{T > t-1\}} \leqslant \big[\psi(\theta)\big]^t. \tag{3.4}$$

For each $i = 1, \dots, d$,

$$\frac{d}{d\theta}\big(\mathbb{E}e^{\theta\langle (\boldsymbol{X}^i_1 - \boldsymbol{e}_i), \boldsymbol{\mu}\rangle}\big)|_{\theta=0} = (\rho-1)\mu_i < x(\rho-1)\mu_{\max} = \frac{d}{d\theta}\big(e^{\theta x(\rho-1)\mu_{\max}}\big)|_{\theta=0},$$

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

since $\rho > 1$ and $\boldsymbol{\mu} > \mathbf{0}$. Because $\psi(0) = 1$, this implies that there exist $\theta_1, \eta_1 > 0$ such that

$$e^{-\eta_1} := \psi(\theta_1)\, e^{-\theta_1 x(\rho-1)\mu_{\max}} < 1.$$

Since the event that $\langle \boldsymbol{S}_t - \boldsymbol{S}_0, \boldsymbol{\mu}\rangle \geqslant x(\rho-1)\mu_{\max}\ t$ implies that $T > t > t-1$, Markov's inequality and inequality (3.4) imply that

$$\begin{aligned}\mathbb{P}\big(\langle \boldsymbol{S}_t - \boldsymbol{S}_0, \boldsymbol{\mu}\rangle \geqslant x(\rho-1)\mu_{\max}\ t\big) &= \mathbb{P}\big(e^{\theta_1\langle \boldsymbol{S}_t-\boldsymbol{S}_0,\boldsymbol{\mu}\rangle}\mathbb{1}_{\{T>t-1\}} \geqslant e^{\theta_1 x(\rho-1)\mu_{\max} t}\big)\\ &\leqslant \mathbb{E}e^{\theta_1\langle \boldsymbol{S}_t-\boldsymbol{S}_0,\boldsymbol{\mu}\rangle}\mathbb{1}_{\{T>t-1\}}e^{-\theta_1 x(\rho-1)\mu_{\max} t}\\ &\leqslant e^{-\eta_1 t}.\end{aligned}$$

So we have proved inequality (3.2). To prove inequality (3.3), we observe that for each $i = 1,\dots,d$

$$\frac{d}{d\theta}\big(\mathbb{E}e^{-\theta\langle (\boldsymbol{X}_1^i-\boldsymbol{e}_i),\boldsymbol{\mu}\rangle}\big)|_{\theta=0} = -(\rho-1)\mu_i < -y(\rho-1)\mu_{\min} = \frac{d}{d\theta}\big(e^{-\theta y(\rho-1)\mu_{\min}}\big)|_{\theta=0},$$

since $\rho > 1$ and $\boldsymbol{\mu} > \mathbf{0}$. Because $\psi(0) = 1$, this implies that there exist $\theta_2, \eta_2 > 0$ such that

$$e^{-\eta_2} := \psi(-\theta_2)\, e^{\theta_2 y(\rho-1)\mu_{\min}} < 1.$$

Since $\mathbb{1}_{\{T>t\}} \leqslant \mathbb{1}_{\{T>t-1\}}$, Markov's inequality and inequality (3.4) imply that

$$\begin{aligned}\mathbb{P}\big(\langle \boldsymbol{S}_t - \boldsymbol{S}_0, \boldsymbol{\mu}\rangle \leqslant y(\rho-1)\mu_{\min} t,\ \|\boldsymbol{S}_t\| > 0\big) &= \mathbb{P}\big(e^{-\theta_2\langle \boldsymbol{S}_t-\boldsymbol{S}_0,\boldsymbol{\mu}\rangle}\mathbb{1}_{\{T>t\}} \geqslant e^{-\theta_2 y(\rho-1)\mu_{\min}\ t}\big)\\ &\leqslant \mathbb{E}e^{-\theta_2\langle \boldsymbol{S}_t-\boldsymbol{S}_0,\boldsymbol{\mu}\rangle}\mathbb{1}_{\{T>t-1\}}e^{\theta_2 y(\rho-1)\mu_{\min} t}\\ &\leqslant e^{-\eta_2 t}.\end{aligned}$$

□

We are now ready to prove the main lemma of this section.

Lemma 3.3. *Let $b > 0$ be constant and set $m = b\log n$. There are constants $\beta, K > 0$ depending on $a_1,\dots,a_d$, λ and b such that if $s = \beta \log n$ then*

$$\mathbb{P}\big(0 < \|\boldsymbol{\mathcal{A}}_s\|_1 \leqslant m\big) = O\big(n^{-(d+1)}\big),$$

and

$$\mathbb{P}\big(\|\boldsymbol{\mathcal{A}}_s\|_1 \geqslant K\log n\big) = O\big(n^{-(d+1)}\big),$$

where $K > b$.

Proof. Our strategy is to prove a slightly stronger statement for $\boldsymbol{S}_t$ (with a lower bound larger than m), then use the coupling between $\boldsymbol{S}_t$ and $\boldsymbol{\mathcal{A}}_t$ in (3.1) to bound the difference between the two. We wish to apply Lemma 3.2 to $\boldsymbol{S}_t$, but the first step of this process is not the same as the subsequent steps, so we must handle this case separately. Notice that in step (1) of the Cluster Discovering Algorithm, the initial vertex does not have a type and is free to search in any of the d directions, while all subsequent vertices are assigned types and thus behave like the multitype branching process with corresponding mean matrix $\boldsymbol{M}_\lambda$. For each $i = 1,\dots,d$, $S_1^{(i)}$ is distributed as a Binomial$(a_i n, \lambda/n)$ random

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

variable. We bound this step of the process as

$$\begin{aligned}\mathbb{P}\big(\langle \boldsymbol{S}_1, \boldsymbol{\mu}\rangle \geqslant (d+2)\log n\big) &= \mathbb{P}\big(e^{\langle \boldsymbol{S}_1,\boldsymbol{\mu}\rangle} \geqslant e^{(d+2)\log n}\big)\\ &\leqslant n^{-(d+2)}\mathbb{E}e^{\langle \boldsymbol{S}_1,\boldsymbol{\mu}\rangle}\\ &\leqslant n^{-(d+2)}\exp\left[\sum_{i=1}^{d} a_i\lambda(e^{\mu_i}-1)\right]\\ &= O\big(n^{-(d+2)}\big). \end{aligned} \tag{3.5}$$

We only need this upper bound, since step (1) of $\boldsymbol{S}_t$ is stochastically bounded below by any of the typical increments (of types $1,\dots,d$), so inequality (3.3) still holds. We now apply Lemma 3.2 to $\boldsymbol{S}_t$ at time $s = \beta\log n$ with $x = 2$ and $y = 1/2$, and use inequality (3.5) to handle step (1) to yield

$$\begin{aligned}\mathbb{P}\big(\langle \boldsymbol{S}_s, \boldsymbol{\mu}\rangle \geqslant \big[2\beta(\rho-1)\mu_{\max} + (d+2)\big]\ \log n\big) &\leqslant n^{-\eta_1\beta} + O\big(n^{-(d+2)}\big)\\ \mathbb{P}\left(\langle \boldsymbol{S}_s, \boldsymbol{\mu}\rangle \leqslant \frac{1}{2}\beta(\rho-1)\mu_{\min}\ \log n,\ \|\boldsymbol{S}_s\| > 0\right) &\leqslant n^{-\eta_2\beta}.\end{aligned}$$

We choose $\beta > 0$ such that $\eta_1\beta > d+1$, $\eta_2\beta > d+1$ and

$$\beta > \frac{2\mu_{\max}(b+1)}{\mu_{\min}(\rho-1)}$$

(recall that $m = b\log n$), and we choose

$$K \geqslant \frac{2\beta(\rho-1)\mu_{\max} + (d+2)}{\mu_{\min}} + \beta$$

such that $K > b$. Using $\langle \boldsymbol{S}_s, \boldsymbol{\mu}\rangle \geqslant \mu_{\min}\|\boldsymbol{S}_s\|_1$ we obtain

$$\mathbb{P}\big(\|\boldsymbol{S}_s\|_1 + s \geqslant K\log n\big) \leqslant n^{-(d+1)}. \tag{3.6}$$

This implies the second part of the lemma, since $\|\boldsymbol{\mathcal{A}}_s\|_1 \leqslant \|\boldsymbol{S}_s\|_1$ with our coupling defined in (3.1). Since $\langle \boldsymbol{S}_s, \boldsymbol{\mu}\rangle \leqslant \mu_{\max}\|\boldsymbol{S}_s\|_1$, we also have

$$\mathbb{P}\big(0 < \|\boldsymbol{S}_s\|_1 \leqslant (b+1)\log n\big) \leqslant n^{-(d+1)}. \tag{3.7}$$

When $\|\boldsymbol{S}_s\|_1 + s \leqslant K\log n$, we have

$$|U_s \cap \mathcal{N}^i(\boldsymbol{v})| \geqslant a_i n - K\log n \tag{3.8}$$

for $\boldsymbol{v}$ of type $j \neq i$. This is because $|A_s| = \|\boldsymbol{\mathcal{A}}_s\|_1 \leqslant \|\boldsymbol{S}_s\|_1$, $|R_s| \leqslant s$ (with equality here if $\boldsymbol{\mathcal{A}}_t$ is still alive at time s), and the maximum number of neighbours that $A_s \cup R_s$ can have in $\mathcal{N}^i(\boldsymbol{v})$ is $|A_s \cup R_s|$ as long as $\boldsymbol{v}$ is not of type i. Recall that $\eta_k^{(t,i)}$ are i.i.d. Bernoulli(λ/n) random variables which are added into the $\boldsymbol{S}_t$ process to account for lost birth opportunities in $\boldsymbol{\mathcal{A}}_t$. Therefore, provided $\boldsymbol{\mathcal{A}}_s > \boldsymbol{0}$, the number of births that occur in $\boldsymbol{S}_t$ which are lost in $\boldsymbol{\mathcal{A}}_t$ for $t \leqslant s$ is at most

$$\sum_{t=0}^{s-1}\sum_{i=1}^{d}\sum_{k=1}^{a_i n - |U_t \cap \mathcal{N}^i(\boldsymbol{v}_t)|} \eta_k^{(t,i)} \leqslant Y^{(s)}, \tag{3.9}$$

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

where

$$Y^{(s)} \sim \text{Binomial}\left(d\beta K(\log n)^2, \frac{\lambda}{n}\right).$$

So, on the event $\|\boldsymbol{S}_s\|_1 + s \leqslant K \log n$, we can bound the difference between $\boldsymbol{S}_s$ and $\boldsymbol{\mathcal{A}}_s$ by a constant, as

$$\begin{aligned}
\mathbb{P}\big(\|\boldsymbol{S}_s\|_1 + s \leqslant K \log n, \|\boldsymbol{\mathcal{A}}_s\|_1 > 0, \|\boldsymbol{S}_s - \boldsymbol{\mathcal{A}}_s\|_1 \geqslant d+2\big) &\leqslant \mathbb{P}\big(Y^{(s)} \geqslant d+2\big) \\
&= \sum_{k=d+2}^{d\beta K(\log n)^2} \binom{d\beta K(\log n)^2}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{d\beta K(\log n)^2 - k} \\
&\leqslant \sum_{k=d+2}^{d\beta K(\log n)^2} \left(\frac{\lambda d\beta K(\log n)^2}{n}\right)^k \\
&= O\big(n^{-(d+1)}\big). \qquad (3.10)
\end{aligned}$$

Finally, combining inequalities (3.6), (3.7), and (3.10) yields

$$\begin{aligned}
\mathbb{P}\big(0 < \|\boldsymbol{\mathcal{A}}_s\|_1 \leqslant (b+1)\log n - (d+2)\big) &= O\big(n^{-(d+1)}\big), \\
\mathbb{P}\big(0 < \|\boldsymbol{\mathcal{A}}_s\|_1 \leqslant b \log n\big) &= O\big(n^{-(d+1)}\big),
\end{aligned}$$

which completes the proof of the lemma. □

This means that, with high probability, at time s the process has either died out or there are at least $m = b \log n$ active vertices in A_s.

3.2. Existence and survival of a lower-bounding random walk

In this section, we will couple the cluster discovering process with a lower-bounding random walk (corresponding to a supercritical branching process) that picks up where the coupling in the last section left off. That is, on the event that the cluster discovering process attains $|A_s| \geqslant m$, we will construct a process $A'_t \subset A_{s+t}$ that is unlikely to die out before some moderately large time $t = r$. To construct this process, we would like to have a lower bound on $|U'_r \cap \mathcal{N}^i(\boldsymbol{v})|$, so the unseen neighbourhood of $\boldsymbol{v}'_t$ at each step in the algorithm will be large, and we will be able to couple a lower-bounding random walk with this process. Using Durrett's discussion of the Erdős–Rényi model as guidance [9], we need $n^{(d-1)/2} \ll r \ll n^{d-1}$, so we choose $r = n^{d-4/3}$.

To generate the lower-bounding process, we begin by letting $p = p_1 + p_2 - p_1 p_2$, where $p_1 = \lambda_1/n$ and $p = \lambda/n$ with $\lambda_1 < \lambda$. Site percolation on $\mathcal{H}$ with parameter p, $\mathcal{H}_p$, can then be viewed as

$$\mathcal{H}_p = \overline{\mathcal{H}_{p_1} \cup \mathcal{H}_{p_2}},$$

where $\mathcal{H}_{p_1}$ and $\mathcal{H}_{p_2}$ are independent site percolations on $\mathcal{H}$, the union is taken over their vertex sets, and the bar denotes the inclusion of all edges from $\mathcal{H}$ between vertices in $\mathcal{H}_{p_1} \cup \mathcal{H}_{p_2}$. It is crucial to note that

$$p_2 = \frac{\lambda - \lambda_1}{n} + \frac{\lambda_1(\lambda - \lambda_1)}{n(n - \lambda_1)} = \frac{\epsilon}{n} + O(n^{-2}),$$

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

where $\epsilon = \lambda - \lambda_1$. The reason for this subdivision is twofold. First, we want to construct a lower-bounding random walk, and for this we will use the parameter p_1. Later, in Section 3.3, we will use a 'sprinkling' argument, which amounts to independently making a small number of additional vertices occupied to connect components of moderate size into a giant component.

For $1 \leqslant k \leqslant d$, let $\boldsymbol{M}_\lambda^{(k)}$ denote the submatrix of $\boldsymbol{M}_\lambda$ consisting of the first k rows and columns, and let $\lambda_c^{(k)}$ be the positive solution to $\det(\boldsymbol{I}^{(k)} - \boldsymbol{M}_\lambda^{(k)}) = 0$. We want to choose $\lambda_1 < \lambda$ so that $\lambda_1 > \lambda_c$ and $\lambda_1 < \lambda_c^{(d-1)}$. In other words, we want λ_1 to be supercritical for the d-dimensional process, but subcritical for the process restricted to any $(d-1)$-dimensional plane. By the monotonicity of the event that a giant component exists, the critical value for the $(d-1)$-dimensional process obtained by considering the first $(d-1)$ dimensions is smallest, since we assumed that $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_d > 0$. The critical value for this process is $\lambda_c^{(d-1)}$, so to guarantee the existence of such a λ_1, we just need to verify that $\lambda_c < \lambda_c^{(d-1)}$. This is easiest to see by using Lemma 1.3. The values of $\det(\boldsymbol{I} - \boldsymbol{M}_{\lambda_c^{(d-1)}})$ and $\det(\boldsymbol{I} - \boldsymbol{M}_0)$ have opposite signs, so by the Intermediate Value Theorem and the uniqueness of λ_c, we have that $0 < \lambda_c < \lambda_c^{(d-1)}$.

We now consider a restricted cluster discovering process that starts with an active set that is a subset of A_s, and proceeds according to the Cluster Discovering Algorithm with vertex probability p_1; call this process (U_t', A_t', R_t') (note that we are resetting time to 0 for the new process). We assume that the initial sets satisfy $(U_0', A_0', R_0') = (U_s, A_s, R_s)$ whenever $|A_s| \leqslant m = b \log n$ (recall that $s = \beta \log n$, and we will decide what b is just after equation (3.18)). If $|A_s| > m$, then we let A_0' consist of the first m vertices of A_s in lexicographic order. The vertices of A_s in excess of m are placed in the removed set, so we let $R_0' = R_s \cup (A_s \setminus A_0')$, and $U_0' = U_s$. Therefore, on the event $|A_s| \geqslant m$, the initial set of active vertices in the process (U_t', A_t', R_t') has size $|A_0'| = m$. Let $\mathbb{P}_{p,p_1}$ denote the probability measure associated with the processes (U_t, A_t, R_t) up to time s with parameter p and (U_t', A_t', R_t') up to time r with parameter p_1.

Let

$$\mathcal{P}_{(i_1,\dots,i_k)}^{(\ell_1,\dots,\ell_k)} := \{\boldsymbol{v} \in V \mid v_{i_1} = \ell_1, \dots, v_{i_k} = \ell_k\} \tag{3.11}$$

denote the intersection of the vertex set, V, with the hyperplanes $v_{i_1} = \ell_1, \dots, v_{i_k} = \ell_k$. We will refer to these as just 'planes', and will emphasize their dimension when it is important to do so. Throughout this argument we will assume that $1 \leqslant i_1, \dots, i_k \leqslant d$, $i_{j_1} \neq i_{j_2}$ for $j_1 \neq j_2$, and $\ell_j \in \{1, \dots, a_{i_j} n\}$, so we have no trivial constraints. We want to bound the number of vertices that become active up to time r in any 2-dimensional plane, $\mathcal{P}_{(i_1,\dots,i_{d-2})}^{(\ell_1,\dots,\ell_{d-2})}$, as this will lead to a lower bound on $|U_t' \cap \mathcal{N}^i(\boldsymbol{v}')|$ for each i and $\boldsymbol{v}'$ that appear in the modified Cluster Discovering Algorithm.

Lemma 3.4. *If $r = n^{d-4/3}$, then there is a constant K_d such that*

$$\begin{aligned}\mathbb{P}_{p,p_1}\Big(&|\mathcal{P}_{(i_1,\dots,i_{d-2})}^{(\ell_1,\dots,\ell_{d-2})} \cap (A_r' \cup R_r')| \geqslant K_d n^{2/3} (\log n)^{d-2} \\ &\qquad \text{for some } (i_1, \dots, i_{d-2}) \text{ and } (\ell_1, \dots, \ell_{d-2})\Big) \\ &= O(n^{-(d+1)}).\end{aligned} \tag{3.12}$$

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

Proof. By Lemma 3.3, with probability $1 - O(n^{-d-1})$, we have that either $|A'_0| = |A_s| = 0$, which implies the complement of the event in (3.12), or $|A'_0| = m$ and $|R'_0| \leqslant s + K \log n = (\beta + K) \log n$. We now only need to consider what happens on the latter event, which we refer to as F.

We begin with $d = 2$. In this case, equation (3.12) is simply the statement that at most $K_2 n^{2/3}$ occupied vertices have been found by time $r = n^{2/3}$, so we merely need a large deviation bound. At each step, we can discover at most a Binomial$(a_1 n, \lambda_1/n)$ number of occupied vertices, so the total number of births up to time r is stochastically bounded by a Binomial$(a_1 nr, \lambda_1/n)$ random variable. By exponentiating and applying Markov's inequality, using the moment generating function for the binomial, and the bound $1 + x \leqslant e^x$, we have

$$\begin{aligned}
&\mathbb{P}_{p,p_1}\left(F \cap \{|A'_r \cup R'_r| \geqslant 2a_1\lambda_1 n^{2/3}\}\right) \\
&\quad \leqslant e^{-2a_1\lambda_1 r}\mathbb{E}_{p,p_1} \exp\left[|A'_r \cup R'_r|\mathbb{1}_F\right] \\
&\quad \leqslant e^{-2a_1\lambda_1 r + (b+\beta+K)\log n}\left[1 + \frac{\lambda_1}{n}(e-1)\right]^{a_1 nr} \\
&\quad \leqslant \exp\left[-2a_1\lambda_1 r + a_1\lambda_1(e-1)r + (b+\beta+K)\log n\right] \\
&\quad = O(n^{-3}),
\end{aligned}$$

where in the last line we used that $e - 1 < 2$. Thus we have (3.12) for $d = 2$ with $K_2 = 2a_1\lambda_1$.

For $d \geqslant 3$, we first consider a fixed $(d-1)$-dimensional plane $\mathcal{P}^{(\ell_1)}_{(i_1)}$. If we restrict our attention to just this plane, ignoring any edges with endpoints outside of $\mathcal{P}^{(\ell_1)}_{(i_1)}$, the resulting subgraph on the vertices in $\mathcal{P}^{(\ell_1)}_{(i_1)}$ will be a subcritical site percolation on a $(d-1)$-dimensional Hamming torus, by the assumptions we made on λ_1. By applying Lemma 2.3 with $\eta = d + 2$, there is a constant β_1 such that with probability at least $(1 - n^{-d-2})$, no vertex in $\mathcal{P}^{(\ell_1)}_{(i_1)}$ will ultimately give rise to a cluster of occupied vertices in $\mathcal{P}^{(\ell_1)}_{(i_1)}$ larger than $\beta_1 \log n$.

If at time t during the Cluster Discovering Algorithm we choose $\boldsymbol{v}'_t \notin \mathcal{P}^{(\ell_1)}_{(i_1)}$, then this vertex may have a single opportunity to discover an occupied vertex in $\mathcal{P}^{(\ell_1)}_{(i_1)}$. We will refer to such an occupied vertex in the focal plane (in this case $\mathcal{P}^{(\ell_1)}_{(i_1)}$) that is discovered by observing the unseen neighbourhood of an occupied vertex that is not in the focal plane as a *seed* vertex. If a potential seed vertex (at the intersection of $\mathcal{P}^{(\ell_1)}_{(i_1)}$ and the line orthogonal to $\mathcal{P}^{(\ell_1)}_{(i_1)}$ that passes through $\boldsymbol{v}'_t$) has already been examined during a previous iteration of the algorithm, then it will not increase the number of seeds in $\mathcal{P}^{(\ell_1)}_{(i_1)}$ whether it is occupied or not. If the potential seed vertex has not been examined (it is in U'_t), then it will be occupied with probability λ_1/n. So, if X_r is the number of seed vertices generated in $\mathcal{P}^{(\ell_1)}_{(i_1)}$ by time r, then X_r is stochastically bounded above by a Binomial$(r, \lambda_1/n)$ random variable. Using this fact and applying a standard generating function argument shows that, for $\gamma_1 = 2\log 2 - 1 > 0$,

$$\mathbb{P}_{p,p_1}\left(X_r \geqslant 2\frac{\lambda_1 r}{n}\right) \leqslant e^{-\gamma_1(\lambda_1 r/n)}.$$

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

Recalling that we have chosen $r = n^{d-4/3}$, and that with probability at least $1 - n^{-d-2}$ each seed gives rise to a cluster in $\mathcal{P}^{(\ell_1)}_{(i_1)}$ of size at most $\beta_1 \log n$, we have

$$\mathbb{P}_{p,p_1}\left(F \cap \{|\mathcal{P}^{(\ell_1)}_{(i_1)} \cap (A'_r \cup R'_r)| \geqslant 2\lambda_1\beta_1 n^{d-7/3}\log n\}\right) \leqslant e^{-\gamma_1\lambda_1 n^{d-(7/3)}} + n^{-d-2}.$$

Since the number of planes, $\mathcal{P}^{(\ell_1)}_{(i_1)}$, is at most $da_1 n$, we have

$$\begin{aligned}\mathbb{P}_{p,p_1}\Big(F \cap \{|\mathcal{P}^{(\ell_1)}_{(i_1)} \cap (A'_r \cup R'_r)| \geqslant 2\lambda_1\beta_1 n^{d-7/3}\log n \text{ for some } (i_1) \text{ and } (\ell_1)\}\Big)\\ = O(n^{-(d+1)}).\end{aligned} \tag{3.13}$$

For the induction step, we wish to show that on the event, E^k_N, that for fixed $k \leqslant d-2$ every $(d-(k-1))$-dimensional plane of the form $\mathcal{P}^{(\ell_1,\ldots,\ell_{k-1})}_{(i_1,\ldots,i_{k-1})}$ has at most N vertices in $A'_r \cup R'_r$, then with probability close to 1, every $(d-k)$-dimensional plane of the form $\mathcal{P}^{(\ell_1,\ldots,\ell_k)}_{(i_1,\ldots,i_k)}$ has at most $2k\lambda_1(N/n)\beta_k \log n$ vertices in $A'_r \cup R'_r$. The reasoning is similar to the argument above for $\mathcal{P}^{(\ell_1)}_{(i_1)}$. If no plane of the form $\mathcal{P}^{(\ell_1,\ldots,\ell_{k-1})}_{(i_1,\ldots,i_{k-1})}$ has more than N vertices in $A'_r \cup R'_r$, then certainly no set of the form

$$\mathcal{P}^{(\ell_1,\ldots,\ell_{k-1})}_{(i_1,\ldots,i_{k-1})} \setminus \mathcal{P}^{(\ell_1,\ldots,\ell_k)}_{(i_1,\ldots,i_k)}$$

has more than N vertices in $A'_r \cup R'_r$. Let us fix $\mathcal{P}^{(\ell_1,\ldots,\ell_k)}_{(i_1,\ldots,i_k)}$, and (allowing for a slight abuse of notation) let

$$\mathcal{P}^{(\ell_1,\ldots,\ell_k)\setminus(\ell_j)}_{(i_1,\ldots,i_k)\setminus(i_j)} := \mathcal{P}^{(\ell_1,\ldots,\ell_{j-1},\ell_{j+1},\ldots,\ell_k)}_{(i_1,\ldots,i_{j-1},i_{j+1},\ldots,i_k)} \setminus \mathcal{P}^{(\ell_1,\ldots,\ell_k)}_{(i_1,\ldots,i_k)}$$

denote the set of vertices inside a $(d-k+1)$-dimensional plane that contains the focal $(d-k)$-dimensional plane, but with the vertices in the focal $(d-k)$-dimensional plane removed. In one iteration of the Cluster Discovering Algorithm, the focal vertex, $\boldsymbol{v}'_t$, is either in $\mathcal{P}^{(\ell_1,\ldots,\ell_k)}_{(i_1,\ldots,i_k)}$, in $\mathcal{P}^{(\ell_1,\ldots,\ell_k)\setminus(\ell_j)}_{(i_1,\ldots,i_k)\setminus(i_j)}$ for some $j = 1,\ldots,k$, or neither. If it is in neither of these sets, then

$$\mathcal{N}(\boldsymbol{v}'_t) \cap \mathcal{P}^{(\ell_1,\ldots,\ell_k)}_{(i_1,\ldots,i_k)} = \emptyset,$$

so this case does not contribute to the number of vertices of $A'_r \cup R'_r$ in the focal plane. If $\boldsymbol{v}'_t \in \mathcal{P}^{(\ell_1,\ldots,\ell_k)}_{(i_1,\ldots,i_k)}$, then by restricting our attention to just this plane as we did above with $\mathcal{P}^{(\ell_1)}_{(i_1)}$, we see that the process looks like a subcritical process in $d-k$ dimensions by our assumptions on λ_1. By Lemma 2.3, there is a constant β_k such that with probability at least $1 - n^{-(d+k+1)}$ all occupied vertices in $\mathcal{P}^{(\ell_1,\ldots,\ell_k)}_{(i_1,\ldots,i_k)}$ belong to clusters restricted to $\mathcal{P}^{(\ell_1,\ldots,\ell_k)}_{(i_1,\ldots,i_k)}$ of size at most $\beta_k \log n$. Finally, if

$$\boldsymbol{v}'_t \in \mathcal{P}^{(\ell_1,\ldots,\ell_k)\setminus(\ell_j)}_{(i_1,\ldots,i_k)\setminus(i_j)} \quad \text{for some } j = 1,\ldots,k,$$

then it has precisely one neighbour in $\mathcal{P}^{(\ell_1,\ldots,\ell_k)}_{(i_1,\ldots,i_k)}$. If this neighbour is unseen (is in U'_t), then it has probability λ_1/n of being an occupied seed vertex. On E^k_N, there are at most kN vertices of $A'_r \cup R'_r$ in

$$\cup_j \mathcal{P}^{(\ell_1,\ldots,\ell_k)\setminus(\ell_j)}_{(i_1,\ldots,i_k)\setminus(i_j)}.$$

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

So if we let Y_k be the number of seed vertices in $\mathcal{P}^{(\ell_1,\dots,\ell_k)}_{(i_1,\dots,i_k)}$, then $\mathbb{1}_{E^k_N} Y_k$ is stochastically bounded above by a Binomial$(kN, \lambda_1/n)$ random variable. Thus, we find that

$$\mathbb{P}_{p,p_1}\left(E^k_N \cap \left\{Y_k \geqslant 2\frac{\lambda_1 kN}{n}\right\}\right) \leqslant e^{-\gamma_1 \frac{\lambda_1 kN}{n}},$$

where $\gamma_1 = 2\log 2 - 1$. Therefore, we have

$$\mathbb{P}_{p,p_1}\left(E^k_N \cap \left\{|\mathcal{P}^{(\ell_1,\dots,\ell_k)}_{(i_1,\dots,i_k)} \cap (A'_r \cup R'_r)| \geqslant 2k\lambda_1\beta_k \frac{N}{n}\log n\right\}\right) \leqslant e^{-\gamma_1\lambda_1 k\frac{N}{n}} + n^{-d-k-1}.$$

Since the number of planes, $\mathcal{P}^{(\ell_1,\dots,\ell_k)}_{(i_1,\dots,i_k)}$, is at most $\binom{d}{k}a_1^k n^k$, provided that $N = \omega(n^{1+\epsilon})$ for some $\epsilon > 0$, we have

$$\begin{aligned}\mathbb{P}_{p,p_1}&\left(E^k_N \cap \left\{\text{some } \mathcal{P}^{(\ell_1,\dots,\ell_k)}_{(i_1,\dots,i_k)} \text{ has } |\mathcal{P}^{(\ell_1,\dots,\ell_k)}_{(i_1,\dots,i_k)} \cap (A'_r \cup R'_r)| \geqslant 2k\lambda_1\beta_k \frac{N}{n}\log n\right\}\right)\\ &= O(n^{-(d+1)}).\end{aligned} \tag{3.14}$$

Finally, combining equations (3.13) and (3.14) to complete the induction argument, we obtain

$$\begin{aligned}\mathbb{P}_{p,p_1}&\left(\text{some } \mathcal{P}^{(\ell_1,\dots,\ell_{d-2})}_{(i_1,\dots,i_{d-2})} \text{ has } |\mathcal{P}^{(\ell_1,\dots,\ell_{d-2})}_{(i_1,\dots,i_{d-2})} \cap (A'_r \cup R'_r)| \geqslant K_d n^{2/3}(\log n)^{d-2}\right)\\ &= O(n^{-(d+1)}),\end{aligned}$$

where

$$K_d = 2^{d-2}\cdot(d-2)!\cdot\lambda_1^{d-2}\cdot\left(\prod_{k=1}^{d-2}\beta_k\right). \qquad \square$$

We now use Lemma 3.4 to get a lower bound on $|U'_t \cap \mathcal{N}^i(\boldsymbol{v}'_t)|$ where $\boldsymbol{v}'_t = (v'_{t1}, v'_{t2}, \dots, v'_{td})$ is a focal vertex of type $j \neq i$ in the cluster discovering process at any time $t \leqslant r$. Each vertex in $\mathcal{P}^{\boldsymbol{j}}_{\boldsymbol{i}} \cap (A'_r \cup R'_r)$, where $\boldsymbol{i} = (i_1, \dots, i_{d-2})$, $i \in \boldsymbol{i}$, and $\boldsymbol{j} = (v'_{ti_1}, \dots, v'_{ti_{d-2}})$, reduces the number of vertices in $U'_t \cap \mathcal{N}^i(\boldsymbol{v}'_t)$ for $t \leqslant r$ by at most one. There are $d-1$ two-dimensional planes of the form $\mathcal{P}^{\boldsymbol{j}}_{\boldsymbol{i}}$, which are parallel to $\boldsymbol{e}_i$ and pass through $\boldsymbol{v}'$. Therefore, Lemma 3.4 implies that with probability at least $1 - O(n^{-(d+1)})$, for $t \leqslant r$ and $\boldsymbol{v}'_t$ of type $j \neq i$,

$$\begin{aligned}|U'_t \cap \mathcal{N}^i(\boldsymbol{v}'_t)| &\geqslant a_i n - (d-1)K_d n^{2/3}(\log n)^{d-2}\\ &\geqslant a_i(1-\delta)n\end{aligned}$$

for sufficiently large n, where $\delta > 0$ is such that $(1-\delta)\lambda_1 > \lambda_c$. This implies that, up to time r, the process (U'_t, A'_t, R'_t) of discovering occupied vertices can be bounded below by the random walk version of a multitype branching process with expectation matrix $\boldsymbol{M}_{(1-\delta)\lambda_1}$, since each focal vertex, $\boldsymbol{v}'_t$, of type j will have an unseen neighbourhood of size at least $a_i(1-\delta)n$ in the $\mathbf{e}_i$ direction for $i \neq j$, and each vertex in this neighbourhood has probability λ_1/n of being occupied. Let us call the random walk version of this multitype branching process $\boldsymbol{W}_t = (W^{(1)}_t, \dots, W^{(d)}_t)$, and assume that $\boldsymbol{W}_0 = \boldsymbol{\mathcal{A}}'_0$. We can couple the

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

processes $\boldsymbol{W}_t$ and $\boldsymbol{\mathcal{A}}'_t$ until time $T_0 \wedge T_\delta \wedge r$, where

$$T_0 := \inf\{t : \boldsymbol{W}_t = \mathbf{0}\},$$
$$T_\delta := \inf\{t : |U'_t \cap \mathcal{N}^i(\boldsymbol{v}'_t)| < a_i(1-\delta)n \text{ for } \boldsymbol{v}'_t \text{ of type } j \text{ and some } i \neq j\}.$$

Note that we have just shown that

$$\mathbb{P}_{p,p_1}\big(T_\delta > r\big) = 1 - O(n^{-(d+1)}). \tag{3.15}$$

We construct the coupling by first choosing $\boldsymbol{v}'_t$ in step (2) of the modified Cluster Discovering Algorithm as follows: first decide that $\boldsymbol{v}'_t$ will be of type i with probability

$$\frac{W_t^{(i)}}{\|\boldsymbol{W}_t\|_1},$$

then choose randomly a vertex from A'_t that is of this type. Let $\mathcal{R}_\delta^{(i)}(\boldsymbol{v}'_t)$ be the unseen neighbourhood of $\boldsymbol{v}'_t$ restricted to the first (lexicographically) $a_i(1-\delta)n$ vertices in $U'_t \cap \mathcal{N}^i(\boldsymbol{v}'_t)$. Then we can complete the coupling of $\boldsymbol{W}_t$ with $\boldsymbol{\mathcal{A}}'_t$ by defining for each $i \in \{1,\dots,d\}$ and $t < T_0 \wedge T_\delta \wedge r$

$$\mathcal{A}'^{(i)}_{t+1} = \mathcal{A}'^{(i)}_t + \mathbb{1}_{\{\boldsymbol{v}'_t \text{ is not of type } i\}} \sum_{\boldsymbol{w} \in U'_t \cap \mathcal{N}^i(\boldsymbol{v}'_t)} \xi'_{\boldsymbol{w}} - \mathbb{1}_{\{\boldsymbol{v}'_t \text{ is of type } i\}}, \tag{3.16}$$

$$W^{(i)}_{t+1} = W^{(i)}_t + \mathbb{1}_{\{\boldsymbol{v}'_t \text{ is not of type } i\}} \sum_{\boldsymbol{w} \in \mathcal{R}_\delta^{(i)}(\boldsymbol{v}'_t)} \xi'_{\boldsymbol{w}} - \mathbb{1}_{\{\boldsymbol{v}'_t \text{ is of type } i\}}, \tag{3.17}$$

where $\xi'_{\boldsymbol{w}}$ is the random variable that takes the value 1 if vertex $\boldsymbol{w}$ is occupied in site percolation with probability parameter p_1, and is 0 otherwise. For $t \geqslant T_0 \wedge T_\delta \wedge r$ we can continue the random walk $\boldsymbol{W}_t$ for all time by defining for each $i \in \{1,\dots,d\}$

$$W^{(i)}_{t+1} = W^{(i)}_t + \sum_{i=1}^{d} \mathbb{1}_{\{J'_{t+1}=i\}}\big(\boldsymbol{X}'_{(t+1,i)} - \mathbf{e}_i\big),$$

where J'_{t+1} are random variables that take the value i with probability

$$\frac{W^{(i)}_t}{\|\boldsymbol{W}_t\|_1},$$

and the random vectors $\boldsymbol{X}'_{(t,i)}$ are independent and equal in distribution to the first step of a multitype branching process started with a single individual of type i, in which a type i individual has a Binomial$(a_j(1-\delta)n, \lambda_1/n)$ number of offspring of type $j \neq i$, and no offspring of type i.

We wish to show now that this coupling will last until time r if $m = |A'_0| = \|\boldsymbol{\mathcal{A}}'_0\|_1$ is large enough. This will occur whenever the branching process corresponding to $\boldsymbol{W}_t$ survives ($T_0 = \infty$) and $T_\delta > r$. Let

$$q'_i = \mathbb{P}\big(\boldsymbol{W}_t = \mathbf{0} \text{ for some } t \mid \boldsymbol{W}_0 = \mathbf{e}_i\big)$$

for $i = 1,\dots,d$. Since the corresponding branching process is supercritical, $q'_i < 1$ for each i. If we let $\gamma_2 = -\log(\max_i q'_i) > 0$, then

$$\mathbb{P}\big(\boldsymbol{W}_t = \mathbf{0} \text{ for some } t \mid \|\boldsymbol{W}_0\|_1 = m\big) \leqslant \big(\max_i q'_i\big)^m = e^{-\gamma_2 m}. \tag{3.18}$$

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

If we let

$$m = \frac{d+1}{\gamma_2}\log n, \quad |A'_0| = m \quad \text{and} \quad |R'_0| = O(\log n),$$

then the coupling between $\boldsymbol{W}_t$ and $\boldsymbol{\mathcal{A}}'_t$ will last until time r with at least probability $1 - n^{-(d+1)}$. Combining equations (3.15) and (3.18) with Lemma 3.3, we have

$$\mathbb{P}_{p,p_1}\big(m \leqslant |A_s| \leqslant K\log n, |A'_r| = 0\big) = O\big(n^{-(d+1)}\big). \tag{3.19}$$

3.3. Merging clusters of size $r = n^{d-4/3}$

At this point, with high probability, the cluster discovering process started at a given vertex, $\boldsymbol{v}$, has either died out by time $s = \beta\log n$ (Lemma 3.3), or will continue to survive until at least time $r = n^{d-4/3}$ (Lemma 3.3, then Lemma 3.4 along with the coupling in equations (3.16)–(3.19)). If the process has died out, then the size of the cluster containing $\boldsymbol{v}$ is at most $\beta\log n$. We will now show that if the site percolation cluster containing the vertex $\boldsymbol{v}_1$ and the site percolation cluster containing the vertex $\boldsymbol{v}_2$ each have at least r vertices, then $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are in the same connected component with probability close to one.

The clusters containing $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ will be generated as follows for $i = 1, 2$.

(1) Start the cluster discovering process, $(U_{t,i}, A_{t,i}, R_{t,i})$, with parameter p at $R_{0,i} = \emptyset$, $A_{0,i} = \{\boldsymbol{v}_i\}$ and $U_{0,i} = V \setminus A_{0,i}$. Continue until time $s = \beta\log n$.
(2) Start the cluster discovering process, $(U'_{t,i}, A'_{t,i}, R'_{t,i})$, with parameter p_1 such that $A'_{0,i}$ consists of the first (at most) m vertices in $A_{s,i}$, $R'_{0,i} = R_{0,i} \cup (A_{s,i} \setminus A'_{0,i})$, and $U'_{0,i} = U_{s,i}$. Continue until time $r = n^{d-4/3}$.
(3) Let $\Upsilon_i = R'_{r,i} \setminus R'_{0,i}$.

Note that the processes for $i = 1$ and $i = 2$ are not independent of one another, but we assume that each set Υ_i is generated without knowledge of the other, so they may or may not have a non-trivial intersection. On the event that each process survives, $|\Upsilon_i| = r$ and $\Upsilon_i \subsetneq C_{\boldsymbol{v}_i}$ for $i = 1, 2$. Also, recall that we had defined the parameter p_2 such that $p = p_1 + p_2 - p_1p_2$, $p = \lambda/n$ and $p_1 = \lambda_1/n$, so $p_2 = \epsilon/n + O(n^{-2})$ where $\epsilon = \lambda - \lambda_1$. The parameter p_2 will be crucial in the proof of the following lemma, which states that if each of the above processes survives, then $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are likely to be in the same component. The proof of the next lemma was simplified significantly due to a suggestion of the anonymous referee, to whom the author is grateful.

Lemma 3.5. *If Υ_1 and Υ_2 are defined for $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ as above, then*

$$\mathbb{P}\big(C_{\boldsymbol{v}_1} \neq C_{\boldsymbol{v}_2}, |\Upsilon_1| = |\Upsilon_2| = r\big) = O(n^{-(d+1)}).$$

Proof. We will employ a sprinkling technique using the probability, p_2, reserved in the first step, where the lower-bounding process was defined using probability parameter p_1. Recall that site percolation on $\mathcal{H}$ with parameter p, $\mathcal{H}_p$, can then be viewed as $\mathcal{H}_p = \overline{\mathcal{H}_{p_1} \cup \mathcal{H}_{p_2}}$, where $\mathcal{H}_{p_1}$ and $\mathcal{H}_{p_2}$ are independent site percolations on $\mathcal{H}$, the union

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

is taken over their vertex sets, and the bar denotes the inclusion of all edges from $\mathcal{H}$ between vertices in $\mathcal{H}_{p_1} \cup \mathcal{H}_{p_2}$.

For the sprinkling argument to work as intended (so we can avoid dependences), we will need to consider only the vertices in $U_{s,1} \cap U_{s,2}$, since the vertices in $V \setminus (U_{s,1} \cap U_{s,2})$ have already been fully considered for inclusion in $\mathcal{H}_p$ in step (1) of the above definition of Υ_1 and Υ_2. However, the number of unseen vertices in any line through an unseen vertex must be large. By Lemma 3.3, with probability $1 - O(n^{-d-1})$ there are at most $(\beta + K)\log n$ vertices in $A_{s,i} \cup R_{s,i}$, and if $\boldsymbol{v} \in U_{s,i}$, then $\boldsymbol{v}$ shares at most $|A_{s,i} \cup R_{s,i}|$ neighbours with $A_{s,i} \cup R_{s,i}$ in each direction. So, if F is the event that $|U_{s,1} \cap U_{s,2} \cap \mathcal{N}^j(\boldsymbol{v})| \geqslant a_j n - 2(\beta + K)\log n$ for all $\boldsymbol{v} \in U_{s,1} \cap U_{s_2}$ and $j \in \{1, \ldots, d\}$, then $\mathbb{P}(F^c) = O(n^{-d-1})$.

Our strategy is to construct a path of occupied vertices in $\mathcal{H}_{p_2}$ from Υ_1 to Υ_2. We begin by looking at the projection of Υ_1 in the $\mathbf{e}_1$ direction onto the coordinate axes, that is,

$$\mathrm{proj}(\Upsilon_1) = \{(x_2, \ldots, x_d) : \exists\ u \text{ s.t. } (u, x_2, \ldots, x_d) \in \Upsilon_1\}.$$

By a standard exponential tail bound for Binomial$(a_i n, p)$ and the union bound,

$$\mathbb{P}\Big(|\mathcal{H}_p \cap \mathcal{P}^{(\ell_1,\ldots,\ell_{d-1})}_{(i_1,\ldots,i_{d-1})}| \geqslant (\log n)^2 \text{ for some } (i_1, \ldots, i_{d-1}) \text{ and } (\ell_1, \ldots, \ell_{d-1})\Big) = O(n^{-\log n + d}),$$

so it is unlikely that any line contains more than $(\log n)^2$ occupied vertices. Since $\mathcal{H}_p$ stochastically dominates Υ_1,

$$\mathbb{P}\big(|\mathrm{proj}(\Upsilon_1)| < r/(\log n)^2, |\Upsilon_1| = r\big) = O(n^{-d-1}).$$

Let E be the event that $|\mathrm{proj}(\Upsilon_1)| \geqslant r/(\log n)^2$.

We now construct paths of occupied vertices starting at vertices in Υ_1 whose projections are distinct elements of $\mathrm{proj}(\Upsilon_1)$; that is, the starting vertices differ in at least one coordinate besides the first. Suppose the first vertex selected is $\boldsymbol{u}^1_0 := (x^1_1, \ldots, x^1_d)$, and we begin by looking for neighbours in the $\mathbf{e}_1$ direction in $\mathcal{H}_{p_2}$. As we search for occupied vertices in $\mathcal{H}_{p_2}$ to construct paths, we will encounter vertices that have already been fully considered for inclusion, either in step (1) of the construction of Υ_i or in the construction of an earlier path. When such a vertex is encountered, we introduce an independent indicator variable that has probability p_2 of being occupied, and we will later bound the probability that such a false vertex was used in a successful path. So, we will find at least one occupied vertex in the $\mathbf{e}_1$ direction with probability

$$1 - (1 - p_2)^{a_1 n} \geqslant 1 - \exp(-\epsilon a_1 + O(n^{-1})) \geqslant \delta_1 > 0$$

for large n, and if there is more than one we choose one uniformly at random and call it $\boldsymbol{u}^1_1 := (u^1_1, x^1_2, \ldots, x^1_d)$. Having found $\boldsymbol{u}^1_i = (u^1_1, \ldots, u^1_i, x^1_{i+1}, \ldots, x^1_d)$, we search for neighbours in the $\mathbf{e}_{i+1}$ direction to obtain $\boldsymbol{u}^1_{i+1} = (u^1_1, \ldots, u^1_{i+1}, x^1_{i+2}, \ldots, x^1_d)$ with probability at least $\delta_{i+1} > 0$. Continuing in this fashion, we ultimately obtain a vertex $\boldsymbol{u}^1_d = (u^1_1, \ldots, u^1_d)$ (with probability larger than $\Delta := \prod \delta_i > 0$), which is uniformly distributed in V. If this experiment is successful (we find a vertex in $\mathcal{H}_{p_2}$ at each step), then we record the path as $\mathrm{path}(1) = \{\boldsymbol{u}^1_0, \ldots, \boldsymbol{u}^1_d\}$ and we call such a path *completed*, otherwise we just record the path up until the point at which we failed to find a vertex.

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

We proceed to construct $n^{1/2}$ paths in the same manner. Each subsequent path starts from a vertex in Υ_1 that projects onto an element of $\mathrm{proj}(\Upsilon_1)$ that has not been used before, and that has not yet been fully considered for inclusion in $\mathcal{H}_p$. We denote the kth path by $\mathrm{path}(k)$ for $1 \leqslant k \leqslant n^{1/2}$, and the ith vertex (if it exists) along the kth path by $\boldsymbol{u}_i^k = (u_1^k, \ldots, u_i^k, x_{i+1}^k, \ldots, x_d^k) \in \mathrm{path}(k)$. Also, we denote the set of all vertices observed through the construction of $\mathrm{path}(k)$ by

$$\mathcal{M}_0 := \emptyset,$$
$$\mathcal{M}_k := \mathcal{M}_{k-1} \cup \bigcup_{i=1}^{d} \mathcal{N}^i\left(\boldsymbol{u}_{i-1}^k\right),$$

where $\mathcal{N}^i(\boldsymbol{u}_{i-1}^k)$ is the set of vertices searched at the ith step of constructing $\mathrm{path}(k)$, and if $\boldsymbol{u}_{i-1}^k$ does not exist, then $\mathcal{N}^i(\boldsymbol{u}_{i-1}^k) = \emptyset$.

The key here is that if $\boldsymbol{u}_d^k \in \mathcal{N}^1(\Upsilon_2)$, then $\mathrm{path}(k)$ connects Υ_1 with Υ_2, and therefore the two components $C_{\boldsymbol{v}_1}$ and $C_{\boldsymbol{v}_2}$ will be joined. If it exists, $\boldsymbol{u}_d^k$ is uniformly distributed in V. Also, by the same argument made above for $\mathrm{proj}(\Upsilon_1)$,

$$\mathbb{P}\left(|\mathcal{N}^1(\Upsilon_2)| < a_1 nr/(\log n)^2, |\Upsilon_2| = r\right) = O(n^{-d-1}).$$

Therefore,

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{u}_d^k \notin \mathcal{N}^1(\Upsilon_2), |\Upsilon_2| = r\right) &\leqslant 1 - \frac{\Delta a_1 nr}{n^d(\log n)^2} + O(n^{-d-1}) \\ &= 1 - \Delta a_1 n^{-1/3}(\log n)^{-2} + O(n^{-d-1}).\end{aligned}$$

Now that we have constructed a collection of paths, $\{\mathrm{path}(k)\}_{k=1}^{n^{1/2}}$, which are conditionally independent given their starting points in Υ_1, we must now bound the probability that $\mathrm{path}(k)$ contains any vertices that were already fully considered for inclusion in $\mathcal{H}_p$. The first possibility is that $\mathrm{path}(k) \cap (V \setminus (U_{s,1} \cup U_{s,2})) \neq \emptyset$. There are not more than $2\beta \log n$ vertices of $\mathrm{proj}(\Upsilon_1)$ that must be avoided, leaving $r/(\log n)^2 - 2\beta \log n$ starting points, and we will only use $n^{1/2}$ of them. On the event F, there are at most $2(\beta + K)\log n$ vertices in

$$\mathcal{N}^i(\boldsymbol{u}_{i-1}^k) \cap (V \setminus (U_{s,1} \cup U_{s,2})),$$

so the probability of including any such vertex in $\mathrm{path}(k)$ is at most

$$1 - [1 - 2(\beta + K)\log n/(a_d n)]^d = O(\log n/n),$$

which bounds the probability that the kth path intersects with $(V \setminus (U_{s,1} \cup U_{s,2}))$.

The second possibility is that $\mathrm{path}(k) \cap \mathcal{M}_{k-1} \neq \emptyset$ for some $k \leqslant n^{1/2}$. Observe that during the discovery of each path we explore at most one neighbourhood in each direction (one $\mathcal{N}^i(\boldsymbol{u}_{i-1}^\ell)$ for each $i = 1, \ldots, d$ and $\ell \leqslant k$) in lexicographic order, and no two starting points $\boldsymbol{u}_0^\ell$ share any neighbours in the $\mathbf{e}_1$ direction. Therefore, the explored neighbourhoods for $\mathrm{path}(k)$ meet the explored neighbourhoods of the first $k-1$ paths orthogonally, if they meet at all, until a collision involving $\mathrm{path}(k)$ occurs, such that $\boldsymbol{u}_i^k \in \mathcal{N}^j(\boldsymbol{u}_{j-1}^\ell)$ for some $i, j \in \{1, \ldots, d\}$ and $\ell < k$. Therefore, until such a collision, each neighbourhood in the discovery of $\mathrm{path}(k)$ sees at most one vertex explored in the construction of $\mathrm{path}(\ell)$ for $\ell < k$. So, if a collision has not occurred by step i of $\mathrm{path}(k)$,

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

then $|\mathcal{M}_{k-1} \cap \mathcal{N}^{i+1}(\boldsymbol{u}_i^k)| \leqslant k-1 \leqslant n^{1/2}$. Therefore, the probability of a collision where $\text{path}(k) \cap \mathcal{M}_{k-1} \neq \emptyset$ is at most $1-[1-n^{1/2}/(a_d n)]^d = O(n^{-1/2})$.

We consider path(k) to be successful if it is completed, $\boldsymbol{u}_d^k$ is in $\mathcal{N}^1(\Upsilon_2)$, and path$(k)$ does not intersect with vertices in $V \setminus (U_{s,1} \cup U_{s,2})$ or $\mathcal{M}_{k-1}$. We have shown that, on the events $E \cap F$ and $|\Upsilon_1| = |\Upsilon_2| = r$ and conditional on the first $k-1 < n^{1/2}$ paths, the probability that path(k) is unsuccessful is at most $1-\Delta a_1 n^{-1/3}(\log n)^{-2} + O(n^{-1/2})$. Therefore, on the events $E \cap F$ and $|\Upsilon_1| = |\Upsilon_2| = r$, the probability of no successful path among the first $n^{1/2}$ paths is at most $[1-\Delta a_1 n^{-1/3}(\log n)^{-2} + O(n^{-1/2})]^{n^{1/2}} \leqslant \exp(-n^{1/7})$ for large n. This bounds the probability that the two components, $C_{\boldsymbol{v}_1}$ and $C_{\boldsymbol{v}_2}$, are disjoint in $\mathcal{H}_p$, and completes the proof of Lemma 3.5. □

3.4. The size of the giant component

To complete the proof of Theorem 1.2, we need to demonstrate that the proportion of occupied vertices included in the giant component approaches $(1-q) > 0$ in probability. To this end we will prove Lemma 3.6, but first we recall the definition of q.

Recall from earlier that $q_i = \mathbb{P}\big(\boldsymbol{Z}_t = \boldsymbol{0} \text{ for some } t \mid \boldsymbol{Z}_0 = \mathbf{e}_i\big)$ is the extinction probability for a multitype branching process in which, for any $k = 1, \ldots, d$, $(Z_1^j \mid \boldsymbol{Z}_0 = \mathbf{e}_k) \sim \text{Poisson}(\lambda a_j)$ if $j \neq k$ and $(Z_1^k \mid \boldsymbol{Z}_0 = \mathbf{e}_k) \equiv 0$, and initially there is one individual of type i. The initial vertex in the cluster discovery process gives birth to $\text{Binomial}(a_i n, \lambda/n)$ neighbours of type i for each i, and henceforth proceeds like the multitype process in which no vertex can give birth to its own type. The limiting branching process is one in which each binomial birth event is replaced with a Poisson birth event with the same mean. If we consider $(d-1)$ independent copies of this Poisson multitype branching process with the modified initial step, and we define q to be the extinction probability of one of these copies, then the collective process will have the same distribution for all time as the multitype branching process that starts with one individual of each of the d types. This implies that $\prod_i q_i = q^{(d-1)}$. From the theory of multitype branching processes [2], the vector $(q_1, \ldots, q_d)$ is the solution to $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}$ for $\boldsymbol{x} \in (0,1)^d$, where

$$f_i(\boldsymbol{x}) = \exp\left[-\lambda \sum_{j \neq i} a_j (1 - x_j)\right].$$

Thus, we have implicitly defined $q < 1$.

Lemma 3.6.

$$\frac{\#\{\boldsymbol{v} \in V : |C_{\boldsymbol{v}}| \leqslant \beta \log n, \xi_{\boldsymbol{v}} = 1\}}{\lambda\big(\prod_i a_i\big) n^{(d-1)}} \longrightarrow q \text{ in probability.} \tag{3.20}$$

The proof of Lemma 3.6 is a fairly standard second-moment argument (a proof of the analogous lemma for the Erdős–Rényi random graph can be found in [9]), so we give only a sketch here.

Proof sketch of Lemma 3.6. The first step is to show that $\mathbb{P}\big(|C_{\boldsymbol{v}}| \leqslant \beta \log n \mid \xi_{\boldsymbol{v}} = 1\big) \to q$. This is easy, since we have already shown that the cluster discovering process, $\boldsymbol{\mathcal{A}}_t$, and

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

the random walk version of the binomial multitype branching process, $\boldsymbol{S}_t$, are identical up to time $s = \beta \log n$ with probability $1 - o(1)$ as per inequality (3.9). Then $\boldsymbol{S}_t$ can be coupled with the corresponding Poisson multitype branching process up to time s, so that the probability that they differ is $o(1)$ [3], which completes this step of the proof.

The next step is to use a second-moment argument to show that the actual proportion of occupied vertices in components smaller than $\beta \log n$ approaches q. To do so, we define the indicator random variables

$$H_{\boldsymbol{v}} = \begin{cases} 1 & \text{if } |C_{\boldsymbol{v}}| \leqslant \beta \log n,\ \xi_{\boldsymbol{v}} = 1, \\ 0 & \text{otherwise,} \end{cases}$$

so that

$$\sum_{\boldsymbol{v} \in V} H_{\boldsymbol{v}} = \#\{\boldsymbol{v} \in V : |C_{\boldsymbol{v}}| \leqslant \beta \log n,\ \xi_{\boldsymbol{v}} = 1\},$$

$$\sum_{\boldsymbol{v} \in V} \mathbb{E} H_{\boldsymbol{v}} = \lambda \left(\prod_i a_i \right) n^{(d-1)}\ \mathbb{P}\big(|C_{\boldsymbol{v}}| \leqslant \beta \log n \mid \xi_{\boldsymbol{v}} = 1\big).$$

If we can show that

$$\lim_{n \to \infty} \frac{\sum_{\boldsymbol{v} \in V} H_{\boldsymbol{v}}}{\sum_{\boldsymbol{v} \in V} \mathbb{E} H_{\boldsymbol{v}}} = 1$$

in probability, then we are done. To apply the second-moment method we need to control the variance of $\sum H_{\boldsymbol{v}}$. We begin by observing that

$$\mathbb{E} H_{\boldsymbol{v}} H_{\boldsymbol{w}} = p^2\ \mathbb{P}\big(|C_{\boldsymbol{v}}| \leqslant \beta \log n,\ |C_{\boldsymbol{w}}| \leqslant \beta \log n \mid \xi_{\boldsymbol{v}} = 1,\ \xi_{\boldsymbol{w}} = 1\big). \tag{3.21}$$

If $\boldsymbol{v}$ and $\boldsymbol{w}$ are not neighbours, then we can first run the cluster discovering process starting at $\boldsymbol{w}$, and check whether $|C_{\boldsymbol{w}}| \leqslant \beta \log n$. If it is, then we also check whether $\boldsymbol{v} \in C_{\boldsymbol{w}}$ (an event referred to as a *collision*). Since $|C_{\boldsymbol{w}}| \leqslant \beta \log n$ and at each step in the cluster discovering process the focal active vertex shares at most two neighbours with $\boldsymbol{v}$, a collision happens with probability at most $1 - (1-p)^{2\beta \log n} = O(n^{-1} \log n)$ on the event that $|C_{\boldsymbol{w}}| \leqslant \beta \log n$. Now, if a collision has not occurred and $|C_{\boldsymbol{w}}| \leqslant \beta \log n$, then we can run the cluster discovering process starting at $\boldsymbol{v}$ and check whether $|C_{\boldsymbol{v}}| \leqslant \beta \log n$. There is a slight wrinkle when doing this, since we must condition on the events that $|C_{\boldsymbol{w}}| \leqslant \beta \log n$ and no collision has occurred. Fortunately, since no collision has occurred, the number of vertices that are observed in each of the $\beta \log n$ steps of the cluster discovering process started at $\boldsymbol{v}$ that were also observed during the cluster discovering process started at $\boldsymbol{w}$ is at most $2\beta \log n$, and all of these vertices must not be occupied. All of the other vertices observed by the cluster discovering process started at $\boldsymbol{v}$ up to time $\beta \log n$ are independent. Therefore, we obtain the following bound on the covariance:

$$\begin{aligned} \mathbb{E} H_{\boldsymbol{v}} H_{\boldsymbol{w}} &\leqslant \frac{\mathbb{E} H_{\boldsymbol{v}} \mathbb{E} H_{\boldsymbol{w}}}{(1-p)^{2\beta^2 (\log n)^2}} + p^2 \mathbb{P}\big(\text{collision}, |C_{\boldsymbol{w}}| \leqslant \beta \log n \mid \xi_{\boldsymbol{v}} = 1,\ \xi_{\boldsymbol{w}} = 1\big) \\ &= (\mathbb{E} H_{\boldsymbol{v}})^2 + O(n^{-3} (\log n)^2). \end{aligned}$$

https://doi.org/10.1017/S096354831300059X Published online by Cambridge University Press

Since $|\mathcal{N}(\boldsymbol{v})| = O(n)$, the contribution to the variance of $\boldsymbol{w} \in \mathcal{N}(\boldsymbol{v})$ is small compared to the mean squared:

$$\sum_{\boldsymbol{v}} \sum_{\boldsymbol{w} \in \mathcal{N}(\boldsymbol{v})} \mathbb{E} H_{\boldsymbol{v}} H_{\boldsymbol{w}} = O(n^d n p^2) = O(n^{d-1}).$$

Therefore, the variance is

$$\operatorname{Var}\left(\sum_{\boldsymbol{v} \in V} H_{\boldsymbol{v}} \right) = O(n^{2d-3} (\log n)^2).$$

So Chebyshev's inequality finishes the proof of Lemma 3.6:

$$\begin{aligned} &\mathbb{P}\left(\left| \frac{\sum_{\boldsymbol{v} \in V} H_{\boldsymbol{v}}}{\sum_{\boldsymbol{v} \in V} \mathbb{E} H_{\boldsymbol{v}}} - 1 \right| > n^{-1/3} \right) \\ &\quad = \mathbb{P}\left(\left| \sum_{\boldsymbol{v} \in V} H_{\boldsymbol{v}} - \sum_{\boldsymbol{v} \in V} \mathbb{E} H_{\boldsymbol{v}} \right| > \lambda \left(\prod_i a_i \right) n^{(d-1)} \, \mathbb{P}\big(|C_{\boldsymbol{v}}| \leqslant \beta \log n \mid \xi_{\boldsymbol{v}} = 1 \big) n^{-1/3} \right) \\ &\quad = O\big(n^{-1/3} (\log n)^2 \big). \end{aligned} \tag{3.22}$$

□

Proof of Theorem 1.2. To summarize, Lemma 3.3 implies that the cluster discovering process started at a vertex $\boldsymbol{v}$ either dies out by time $\beta \log n$ or has at least $m = (d+1)/\gamma_2 \log n$ active vertices at this time. If the process survives to time $\beta \log n$, then Lemma 3.4 implies that the cluster discovering process dominates a supercritical branching process, so we get equation (3.19), which says that the process will not die out before time $r = n^{d-4/3}$. So with high probability, every occupied vertex is either in a component of size at most $\beta \log n$ or at least $n^{d-4/3}$. Lemma 3.5 implies that all vertices in components of size at least $n^{d-4/3}$ are actually in a single large component. Finally, Lemma 3.6 says that the proportion of occupied vertices that are in small components is asymptotically q, so the proportion in the giant component is $(1-q)$. □

Acknowledgements

The author thanks Janko Gravner for many helpful discussions. The author is grateful to the anonymous referee whose comments greatly improved the content and presentation of this article. In particular, a suggestion of the referee led to a substantially simplified proof of Lemma 3.5 that appears here.

References

[1] Ajtai, M., Komlós, J. and Szemerédi, E. (1981) Largest random component of a k-cube. *Combinatorica* **2** 1–7.
[2] Athreya, K. B. and Ney, P. E. (1972) *Branching Processes*, Dover.
[3] Barbour, A. D., Holst, L. and Janson, S. (1992) *Poisson Approximation*, Oxford University Press.
[4] Bollobás, B., Janson, S. and Riordan, O. (2007) The phase transition in inhomogeneous random graphs. *Random Struct. Alg.* **31** 3–122.

[5] Bollobás, B., Janson, S. and Riordan, O. (2009) Line-of-sight percolation. *Combin. Probab. Comput*. **18** 83–106.
[6] Bollobás, B., Kohayakawa, Y. and Łuczak, T. (1991) On the evolution of random boolean functions. In *Extremal Problems for Finite Sets, Visegrád, Hungary*, pp. 137–156.
[7] 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 Struct. Alg*. **27** 137–184.
[8] 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.
[9] Durrett, R. (2007) *Random Graph Dynamics*, Cambridge University Press.
[10] Erdős, P. and Rényi, A. (1960) On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*. **5** 17–61.
[11] van der Hofstad, R. and Luczak, M. (2009) Random subgraphs of the 2D Hamming graph: The supercritical phase. *Probab. Theory Rel. Fields* **147** 1–41.