J. Appl. Prob. **54,** 1278–1294 (2017)
*doi:*10.1017/jpr.2017.63

PHASE TRANSITION IN RANDOM DISTANCE GRAPHS ON THE TORUS

FIORALBA AJAZI,* *Lund University and University of Lausanne*
GEORGE M. NAPOLITANO ** AND
TATYANA TUROVA,* *** *Lund University*

Abstract

In this paper we consider random distance graphs motivated by applications in neurobiology. These models can be viewed as examples of inhomogeneous random graphs, notably outside of the so-called rank-1 case. Treating these models in the context of the general theory of inhomogeneous graphs helps us to derive the asymptotics for the size of the largest connected component. In particular, we show that certain random distance graphs behave exactly as the classical Erdős–Rényi model, not only in the supercritical phase (as already known) but in the subcritical case as well.

Keywords: Inhomogeneous random graph; random distance graph; largest connected component

2010 Mathematics Subject Classification: Primary 05C80; 60J85

1. Introduction

Random distance graphs are often designed as models of real-world systems where some of the properties of the connections between vertices are observed to be dependent on their relative distance (for some early examples, see, e.g. [1], [3], and [15]). It is generally assumed that the vertices of such models are in some metric space, most often $\mathbb{R}^d$ or $\mathbb{Z}^d$. The probability of a connection between any two vertices in these graphs is a function of the distance between them.

A great demand for this class of models is prompted in particular by developments in neuroscience. It must be noted that physiological data on the brain structure of a living organism is a highly costly exercise (see, e.g. the Blue Brain Project of [12]), hence, inevitably it has to be complemented with theoretical studies. A number of models have been developed along this line to make mathematical results accessible for applications and, in particular, random graphs have become a common tool in the exploration of neuronal networks (see, e.g. [16] and [18] and the references therein).

Recently, a model for the structure and dynamics of the neuropil has been proposed by Janson *et al.* [10]. Inspired by this work, we consider here certain random distance graphs, whose vertices lie on a two-dimensional discrete torus and the connection probabilities decay both with the distance between the vertices and the total number of vertices in the graph (see Section 2 for a precise definition). Our results deal with one of the primary questions, namely,

Received 6 December 2016; revision received 14 August 2017.
* Postal address: Department of Mathematical Statistics, Faculty of Science, Lund University, Sölvegatan 18, 22100 Lund, Sweden.
** Current address: Centre for Epidemiology and Screening, Department of Public Health, University of Copenhagen, Øster Farimagsgade 5, 1014 Copenhagen, Denmark.
*** Email address: tatyana@maths.lth.se

the size of the largest connected component. This characteristic is very important for the study of processes on the networks, as, e.g. the propagation of impulses. In complex dynamical systems the parameters of connectivity change in time; for neuronal networks this is a known property of the synaptic plasticity. Therefore, it is important to have a complete picture of the scalings for the largest connected component on the entire parameter space, at least for some basic test networks. Such a complete description should help to fit the parameters of the connections in a neuronal model based only on qualitative information on the functioning of a network.

As a mathematical object, random distance graphs form a particular subclass of the general inhomogeneous random graph models [6]. The graphs treated in [6] have the following common feature: edges are placed independently from each other and the probability of edges is, roughly speaking, of order $1/n$, where n is the size of the graph (i.e. the number of its vertices). Briefly, each of the n vertices is assigned a type, i.e. a value in some separable metric space $\mathcal{S}$. Given a set of such values $\{x_1, \dots, x_n\}$ any two vertices i and j are connected with probability

$$p_n(i, j) = \min\left\{\frac{\kappa(x_i, x_j)}{n}, 1\right\}, \tag{1}$$

where κ is a symmetric nonnegative measurable function.

Most investigations on random distance graphs have been carried out without much use of [6] (not counting Example 4.6 of [6] itself, Bollobás *et al.* [7] is almost an exception). The reason, perhaps, is that random distance graphs are outside of the so-called rank-1 case, and thus they belong to a complicated subclass of the inhomogeneous models. The theory of [6] gives us the critical parameters for the emergence of the giant component and even describes the size of this component in the supercritical phase. However, the subcritical phase of non rank-1 models was studied only for some particular subclasses (see [19]), which do not include the present model. Furthermore, the critical phase has been studied so far only for the rank-1 cases (see [4], [5], and [20]).

The paper is organized as follows. In Section 2 we define our model and outline the connections with some random graphs models previously studied. The main theorems are stated in Section 3, whereas their proofs are collected in Section 4.

2. The model

Let $N \in \mathbb{N}$, $N > 1$, and let $V_N = \{1, \dots, N\}^2$ denote the set of vertices in the two-dimensional discrete torus $\mathbb{T}_N^2 = (\mathbb{Z}/N\mathbb{Z})^2$. Define the graph distance $d(u, v)$ between two vertices $u = (u_1, u_2)$ and $v = (v_1, v_2)$ in V_N as

$$d(u, v) = d_N(|u_1 - v_1|) + d_N(|u_2 - v_2|), \tag{2}$$

where

$$d_N(i) = \begin{cases} i, & 0 \le i \le N/2, \\ N - i, & N/2 < i < N, \end{cases} \quad \text{for } i \in \{0, \dots, N-1\}.$$

Let W be a nonnegative random variable, and let W_v, $v \in V_N$, be independent and identically distributed (i.i.d.) copies of W. Given the values W_v, $v \in V_N$, assume that between any two vertices $u, v \in V_N$ an edge is present independently of others and with probability

$$p(u, v) = \min\left\{c\frac{W_u W_v}{N d(u, v)}, 1\right\}, \tag{3}$$

where $c > 0$ is a parameter. We denote by $G_{N,W}$ the resulting random graph on V_N.

https://doi.org/10.1017/jpr.2017.63 Published online by Cambridge University Press

Note that, in the case of constant $W \equiv 1$, this graph is exactly the one introduced by Janson *et al.* [10] and it also has common features with other random graph models considered previously (see, e.g. [18]). Janson *et al.* [10] studied a bootstrap percolation process as a model of the spread of activation in a neuronal tissue. They also derived the size of the diameter of the graph, thus extending the corresponding results of [6] for graphs with unbounded number of types.

In the language of inhomogeneous graph theory, the model introduced by Janson *et al.* [10] can be seen as a *homogeneous* case (that is, roughly speaking, when the degrees of the vertices are asymptotically all the same; see Example 4.6 of [6]). Note, however, that the general form (3) considered here (inspired by [8]) makes the model essentially inhomogeneous.

It is worth mentioning that the model (3) is also closely related to certain bond percolation models (see [15] and, in particular, [1] and [14]). In such models, the graphs have a countable set of vertices and, as in the model we investigate here, edges between them are present with a distance-dependent probability. The main problem is whether, depending on the values of the parameters, a particular vertex belongs to an infinite cluster with positive probability. It was shown in [7] that this question for the spread-out percolation model [14] can be resolved using the theory of inhomogeneous random graphs.

3. Results

It has been already shown in [6, Example 4.6] that in the supercritical case a homogeneous distance graph has the same asymptotics for the size of largest connected component as in the classical Erdős–Rényi model. We prove that this result holds for the subcritical case as well.

Theorem 1. *Let G_N denote a random graph on V_N with probability of connections*

$$p(u, v) = \min\left\{\frac{c}{Nd(u, v)}, 1\right\}, \qquad u, v \in V_N,$$

and let $C(G_N)$ denote the size of the largest connected component in G_N. Set

$$\lambda = c4\log 2.$$

Then the following hold.

(i) *If $\lambda < 1$, we have*

$$\frac{C(G_N)}{\log(N^2)} \xrightarrow{\mathbb{P}} \frac{1}{\lambda - 1 - \log\lambda} \quad \text{as } N \to \infty. \tag{4}$$

(ii) *If $\lambda > 1$ then*

$$\frac{C(G_N)}{N^2} \xrightarrow{\mathbb{P}} \beta \quad \text{as } N \to \infty,$$

where $\beta = \beta(\lambda)$ is the positive solution of $\beta = 1 - \mathrm{e}^{\lambda\beta}$.

As we noted above, only Theorem 1(ii) follows from the results of [6].

Remark 1. One may observe a certain redundancy here, as statements (i) and (ii) of Theorem 1 are particular cases of the following Theorems 3 and 2, respectively. However, stated separately, Theorem 1 makes it clear that the largest connected component in G_N behaves, asymptotically, as the one in the Erdős–Rényi graph $G_{n,p}$, with $n = N^2$ and $p = \lambda/n$.

Furthermore, it is plausible to conjecture (but we do not study this case here) that the analysis of the critical phase in [20] can be extended to this model as well. This would yield that even

https://doi.org/10.1017/jpr.2017.63 Published online by Cambridge University Press

in the critical case, that is, when $\lambda = c4\log 2 = 1$, the graph G_N has the same asymptotics for the largest component as $G_{n,p}$, with $p = 1/n$ and $n = N^2$, i.e. that the largest connected component rescaled by $n^{2/3}$ converges in distribution to a certain positive random variable.

The following theorems treat the general model (3).

Theorem 2. *Assume that*

$$\mathbb{E}W^2 = \int_0^\infty x^2 \mu_W(\mathrm{d}x) < \infty.$$

Let $C(G_{N,W})$ denote the size of the largest connected component in $G_{N,W}$, and denote again

$$\lambda = c4\log 2.$$

Then

$$\frac{C(G_{N,W})}{N^2} \xrightarrow{\mathbb{P}} \int_0^\infty \beta(x)\mu_W(\mathrm{d}x) =: \hat{\beta}, \tag{5}$$

where $\beta(x)$ is the maximal solution to

$$f(x) = 1 - \exp\left(x\lambda \int_0^\infty y f(y)\mu_W(\mathrm{d}y)\right). \tag{6}$$

Furthermore, $\hat{\beta} > 0$ if and only if

$$\lambda\mathbb{E}W^2 > 1. \tag{7}$$

Note that the critical parameter $\lambda\mathbb{E}W^2$ in Theorem 2 is similar to the lower bound derived in Theorem 4.1 of [8] (in fact, it has exactly the same meaning of a certain averaged degree of a vertex as in [8]) to provide the necessary conditions for percolation.

Theorem 2 follows essentially from the general theory of [6], as we explain below. It tells us that the limit when $N \to \infty$ of the (scaled) largest component in $G_{N,W}$ coincides with the corresponding limit for the rank-1 random graph on V_N defined by the following probabilities of connections between any $u, v \in V_N$:

$$p_1(u, v) = \min\left\{\lambda \frac{W_u W_v}{N^2}, 1\right\}. \tag{8}$$

(Note, however, that for any finite N, models (3) and (8) are not equal in distribution.)

Here the largest connected component in the subcritical phase is sensitive to the tail of the distribution of W. It is known that in models of the form in (8), the size of the largest component varies between polynomial (see [9]) and logarithmic (see [17]) order depending on the distribution of W. We shall consider here a particular case of the distribution of W to show the similarities with Theorem 1.

Theorem 3. *Assume that, for some positive ε, $\mathbb{E}\mathrm{e}^{\varepsilon W} < \infty$. If also $\lambda\mathbb{E}W^2 < 1$, there exists a unique $y > 1$ which satisfies*

$$y = \frac{1}{\lambda M} \frac{\mathbb{E}(W\mathrm{e}^{\lambda M(y-1)W})}{\mathbb{E}(W^2\mathrm{e}^{\lambda M(y-1)W})},$$

where $M = \mathbb{E}W$. Let $C(G_{N,W})$ be the size of the largest connected component in $G_{N,W}$. Then we have

$$\frac{C(G_{N,W})}{\log(N^2)} \xrightarrow{\mathbb{P}} \frac{1}{\log\gamma} \quad \text{as } N \to \infty,$$

where

$$\gamma := \frac{1}{\lambda\mathbb{E}(W^2\mathrm{e}^{\lambda M(y-1)W})} > 1.$$

https://doi.org/10.1017/jpr.2017.63 Published online by Cambridge University Press

Finally, we remark that our analysis based on the inhomogeneous random graph theory is well applicable for models of type (3) even for different distance functions d as, e.g. euclidean distance, as long as one can justify the relations similar to (9) and (27). Also, our approach can be generalized to the similar models in higher dimensions.

4. Proofs

4.1. Random distance graph via inhomogeneous random graphs

Rescale the set V_N as follows:

$$V_N \to \widetilde{V}_N = \left\{ \left(\frac{u_1}{N}, \frac{u_2}{N} \right) \colon (u_1, u_2) \in V_N \right\}.$$

Hence, $\widetilde{V}_N$ is a set of N^2 vertices in the continuous torus $\mathbb{T}^2 := (\mathbb{R}/\mathbb{Z})^2$. Let $\mu_{\mathcal{L}}$ denote the Lebesgue measure on $\mathbb{T}^2$, and let μ_W be the Borel measure on $\mathbb{R}_+$ induced by the random variable W. Denote $\mathcal{S} := \mathbb{T}^2 \times \mathbb{R}_+$, and define the product measure $\mu = \mu_{\mathcal{L}} \times \mu_W$ on this space. Then the triple $\mathcal{V} := (\mathcal{S}, \mu, \{(v, W_v) \colon v \in \widetilde{V}_N\})$ satisfies the definition of a *generalized vertex space* from [6], i.e. for any Borel set $A \subseteq \mathcal{S}$, the following convergence holds:

$$\frac{|\{v \colon (v, W_v) \in A\}|}{N^2} \xrightarrow{\mathbb{P}} \mu(A).$$

Define now, for $u \neq v,\ u, v \in \mathbb{T}^2$, the kernel

$$\kappa_1(u, v) := \frac{1}{\rho(u, v)},$$

where, for any $u = (u_1, u_2)$, $v = (v_1, v_2) \in \mathbb{T}^2$,

$$\rho(u, v) = \rho_1(|u_1 - v_1|) + \rho_1(|u_2 - v_2|)$$

with

$$\rho_1(a) = \begin{cases} a, & 0 \le a \le \frac{1}{2}, \\ 1 - a, & \frac{1}{2} < a \le 1. \end{cases}$$

Furthermore, let $\kappa_2(x, y) := xy$ denote the standard product on $\mathbb{R}_+^2$.

Finally, we define the kernel on $\mathcal{S} \times \mathcal{S}$:

$$\kappa((u, x), (v, y)) := \kappa_1(u, v)\kappa_2(x, y), \qquad (u, x), (v, y) \in \mathcal{S},$$

and construct the random graph $G^{\mathcal{V}}(N^2, \kappa)$ (following the notation of [6]) on a given set $\widetilde{V}_N$ of N^2 vertices in $\mathcal{S}$, by placing an independent edge between any pair of vertices $\boldsymbol{x}_i, \boldsymbol{x}_j \in \widetilde{V}_N$ with probability (see (1))

$$\widetilde{p}(\boldsymbol{x}_i, \boldsymbol{x}_j) := \min\left\{ c\frac{\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)}{N^2}, 1 \right\}.$$

Proposition 1. *The model G_N is equivalent to the inhomogeneous random graph model $G^{\mathcal{V}}(N^2, \kappa)$.*

Proof. Given a set of types $w_v,\ v \in V_N$, let $\widetilde{v} = v/N$ for any $v \in V_N$. Then the probability of connection (3) satisfies

$$p(u, v) = \min\left\{ c\frac{w_u w_v}{Nd(u, v)}, 1 \right\} = \min\left\{ c\frac{\kappa_1(\widetilde{u}, \widetilde{v}) w_u w_v}{N^2}, 1 \right\} = \widetilde{p}((\widetilde{v}, w_v), (\widetilde{u}, w_u)). \qquad (9)$$

https://doi.org/10.1017/jpr.2017.63 Published online by Cambridge University Press

Hence, given a set of types w_v, $v \in V_N$, there is a connection between any two vertices $u, v \in V_N$ of G_N if and only if there is a connection between the corresponding vertices $(\widetilde{u}, w_u)$ and $(\widetilde{v}, w_v)$ of the graph model $G^{\mathcal{V}}(N^2, \kappa)$. □

It is straightforward to check that the kernel κ is *graphical* (see Definition 2.7 of [6]), since it is continuous, $\kappa \in L^1(\mathcal{S} \times \mathcal{S}, \mu \times \mu)$, and the number of edges in the graph $e(G^{\mathcal{V}}(N^2, \kappa))$ satisfies the following convergence:

$$\frac{1}{N^2}\mathbb{E}e(G^{\mathcal{V}}(N^2, \kappa)) \to \frac{1}{2}\int_{\mathcal{S}}\int_{\mathcal{S}} \kappa(\boldsymbol{x}, \boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}. \tag{10}$$

4.2. Proof of Theorem 2

Proposition 1, together with (10), allows us to apply some of the results of [6] to our case. In particular, we can approximate the size of the connected component by the total progeny of a multitype Galton–Watson branching process $\mathcal{B}(\kappa)$, with type-space $\mathcal{S}$, where the single ancestor has type $\boldsymbol{x}$, and the number of offspring of type $\boldsymbol{y}$ of each individual of type $\boldsymbol{x} \in \mathcal{S}$ has Poisson distribution with intensity $\kappa(\boldsymbol{x}, \boldsymbol{y})\mu(\mathrm{d}\boldsymbol{y})$. Denote here $\beta_\kappa(\boldsymbol{x})$ and $\mathcal{X}(\boldsymbol{x})$, correspondingly, the survival probability and the size of the total progeny of this branching process with the ancestor of type $\boldsymbol{x}$.

Following [6], let us define the integral operator T_κ:

$$(T_\kappa f)(\boldsymbol{x}) := \int_{\mathcal{S}} \kappa(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y})\,\mathrm{d}\mu(\boldsymbol{y})$$

for all measurable functions f (when the integral is defined) on $\mathcal{S}$, and define the norm of T_κ as

$$\|T_\kappa\| := \sup\{\|T_\kappa f\|_2 \colon f \geq 0,\ \|f\|_2 \leq 1\}. \tag{11}$$

Then, by Theorem 3.1 of [6] (whose applicability here is granted by Proposition 1), we immediately obtain

$$\frac{C(G_{N,W})}{N^2} \xrightarrow{\mathbb{P}} \int_{\mathcal{S}} \beta_\kappa(\boldsymbol{x})\mu(\mathrm{d}\boldsymbol{x}) =: \hat{\beta}. \tag{12}$$

Moreover, it was also proved in [6] that the survival probability β_κ is the maximal solution to

$$f(\boldsymbol{x}) = 1 - \mathrm{e}^{-(T_\kappa f)(\boldsymbol{x})}, \qquad \boldsymbol{x} \in \mathcal{S}, \tag{13}$$

and that $\hat{\beta} > 0$ if and only if

$$\|T_\kappa\| > 1. \tag{14}$$

Observe that it follows directly from the symmetry of our model that the survival probability $\beta_\kappa(\boldsymbol{x})$, where $\boldsymbol{x} = (u, x) \in \mathbb{T}^2 \times \mathbb{R}_+$, does not depend on $u \in \mathbb{T}^2$, but it is only a function of $x \in \mathbb{R}_+$. Hence, we shall simply write the survival probability as

$$\beta_\kappa(\boldsymbol{x}) = \beta_\kappa(x), \qquad \boldsymbol{x} = (u, x) \in \mathcal{S},$$

which, by (13), is the maximal solution to

$$f(x) = 1 - \exp\left(-\lambda \int_0^\infty xyf(y)\mu_W(\mathrm{d}y)\right), \qquad x \in \mathbb{R}_+, \tag{15}$$

i.e. (6). This together with (12) yields (5).

https://doi.org/10.1017/jpr.2017.63 Published online by Cambridge University Press

We are left to prove (7). Firstly, one could use definition (11) to derive straightforwardly

$$\|T_\kappa\| = \lambda \mathbb{E} W^2,$$

which together with (14) would yield (7). However, it is easier to derive (7) using direct relations between the defined above multitype branching process and a certain homogeneous Galton–Watson process which we define shortly.

Let us introduce first yet another branching process $\mathcal{B}_1(x)$ with type-space $\mathbb{R}_+$, where the single ancestor has type x, and the number of offspring of type $y \in \mathbb{R}_+$ of any individual of type $x \in \mathbb{R}_+$ has Poisson distribution with intensity $\lambda x y \mu_W(\mathrm{d}y)$. Using the same analysis as for $\mathcal{B}$, we obtain the survival probability $\beta^1_\kappa(x)$ of $\mathcal{B}_1(x)$ as the maximum solution to (15). Therefore, in the notation of Theorem 2, it holds that

$$\beta(x) = \beta^1_\kappa(x) = \beta_\kappa(\boldsymbol{x})$$

for all $u \in \mathbb{T}^2$ and for any $\boldsymbol{x} = (u, x)$.

Finally, we define a homogeneous Galton–Watson process $\mathcal{B}_2$. This process starts with one single ancestor and its offspring distribution $\tilde{Y}$ has a compound Poisson distribution

$$\mathrm{Poisson}(\widetilde{W}\lambda\mathbb{E}(W)),$$

where the random variable $\widetilde{W}$ has the following so-called size-biased distribution:

$$\mu_{\widetilde{W}}(\mathrm{d}y) := \frac{y\mu_W(\mathrm{d}y)}{\mathbb{E}W}.$$

Let us denote $\mathcal{X}_1(x)$ and $\mathcal{X}_2$ as the total progeny of $\mathcal{B}_1(x)$ and $\mathcal{B}_2$, respectively. It was proved in Section 2.2 of [17], that $\mathcal{X}_1(\widetilde{W})$ and $\mathcal{X}_2$ are equal in distribution, i.e.

$$\mathcal{X}_1(\widetilde{W}) \overset{\mathrm{D}}{=} \mathcal{X}_2. \tag{16}$$

In the case of a homogeneous process $\mathcal{B}_2$, the necessary and sufficient condition for a positive survival probability is simply $\mathbb{E}(\tilde{Y}) = \lambda\mathbb{E}(W^2) > 1$. Therefore, (16) yields $\mathbb{P}(\widetilde{\mathcal{X}}_1(\widetilde{W}) = \infty) > 0$ if and only if $\lambda\mathbb{E}(W^2) > 1$.

It follows by the properties of a Poisson distribution that the type of a randomly chosen offspring of the ancestor in the process $\mathcal{B}_1(x)$ has distribution $\widetilde{W}$ for any $x \in \mathbb{R}_+$. Hence, for any x, the process $\mathcal{B}_1(x)$ survives with a positive probability (i.e. $\beta^1_\kappa(x) > 0$), if $\mathcal{B}_1(\widetilde{W})$ survives with a positive probability (i.e. $\mathbb{P}(\widetilde{\mathcal{X}}_1(\widetilde{W}) = \infty) > 0$). Since β^1_κ is the maximal solution to (15), i.e. to (6), it follows that $\hat{\beta} > 0$ (see (5)) if $\lambda\mathbb{E}(W^2) > 1$.

On the other hand, if $\mathcal{B}_1(\widetilde{W})$ survives with probability 0 (i.e. if $\lambda\mathbb{E}(W^2) \leq 1$) then the equality

$$0 = \mathbb{P}(\widetilde{\mathcal{X}}_1(\widetilde{W}) = \infty) = \int_{\mathbb{R}_+} \mathbb{P}(\widetilde{\mathcal{X}}_1(x) = \infty)\mu_{\widetilde{W}}(\mathrm{d}x) = \int_{\mathbb{R}_+} \beta^1_\kappa(x)\mu_{\widetilde{W}}(\mathrm{d}x)$$

implies that $\beta^1_\kappa = 0$, $\mu_{\widetilde{W}}$-almost surely (a.s.), and, hence, μ_W-a.s. Since β^1_κ is the maximal solution to (15), i.e. to (6), it follows that in this case $\hat{\beta} = 0$.

Summarizing, we find that $\hat{\beta} > 0$ if and only if $\lambda\mathbb{E}(W^2) > 1$. In turn, this yields $\|T_\kappa\| = \lambda\mathbb{E}(W^2)$. This proves the theorem. □

4.3. Proof of Theorem 1

4.3.1. *Breadth-first search.* Let us fix a vertex $v \in V_N$ arbitrarily and let $C_v(N)$ denote the connected component containing v. We use a standard procedure to reveal $C_v(N)$, an exploration algorithm known as the *breadth-first search* (see, e.g. [2] or [21]). This is defined as follows.

https://doi.org/10.1017/jpr.2017.63 Published online by Cambridge University Press

In the course of exploration, the vertices of G_N can be in one of the three states: *active*, *saturated*, or *neutral*. At time $i = 0$, the vertex v is set to be active, while all the other vertices are neutral. This ends step $i = 0$.

We denote by S_i the set of active vertices at time i. Hence, $|S_0| = 1$. The state of a vertex changes during the exploration of $C_v(N)$ as follows.

At each time step $i \geq 1$, we choose an active vertex in S_{i-1} uniformly at random and denote it by v_i. Then each vertex u which is neutral after step $i-1$ becomes active at step i, if it is connected to v_i; otherwise, u stays neutral. After searching the entire set of neutral vertices the vertex v_i becomes saturated. This finishes the ith step of the exploration algorithm.

The process stops when there are no more active vertices, i.e. at the first time i when $|S_i| = 0$, that is, at time

$$T = \min\{i \geq 1 \colon |S_i| = 0\}. \tag{17}$$

At this time all considered vertices are saturated and they do not have any connection to the neutral vertices. Hence, $C_v(N)$ coincides with the set of saturated vertices, and, thus, $|C_v(N)| = T$.

Let X_i denote the number of vertices becoming active at the ith step. Then the following recursion holds:

$$|S_0| = 1, \qquad |S_i| = |S_{i-1}| + X_i - 1 = X_1 + \cdots + X_i - (i-1). \tag{18}$$

Correspondingly, we can rewrite T, defined in (17), as follows:

$$T = \min\{i \geq 1 \colon X_1 + \cdots + X_i = i - 1\}.$$

4.3.2. *Subcritical case.* In this section we assume that $\lambda < 1$ and we prove part (i) of Theorem 1.

Upper bound. We start by finding an upper bound on $C(G_N)$, the size of the largest connected component. Namely, we prove that, for any positive ε,

$$\mathbb{P}\left\{\frac{C(G_N)}{\log N} < \frac{2}{1-\lambda-\log\lambda} + \varepsilon\right\} \to 1 \quad \text{as } N \to \infty. \tag{19}$$

The proof is based on the exploration algorithm described above. We also use essentially the geometry of the discrete torus with the distance defined in (2). Recall, in particular, that the number N_r of vertices at distance r from any given vertex, for N odd, is given by

$$N_r = \begin{cases} 4r, & 1 \leq r \leq \lfloor N/2 \rfloor, \\ 4(N-r), & \lfloor N/2 \rfloor < r \leq N; \end{cases} \tag{20}$$

while for N even, it is given by

$$N_r = \begin{cases} 4r, & 1 \leq r < N/2, \\ 2(N-1), & r = N/2, \\ 4(N-r), & N/2 < r < N, \\ 1, & r = N. \end{cases} \tag{21}$$

Recall that the vertices becoming active at the ith step are connected to the vertex v_i. Let $X_{i,r}$ denote the number of vertices at distance r from vertex v_i, which become active at the ith step. Hence,

$$X_i = \sum_{r=1}^{N} X_{i,r}.$$

https://doi.org/10.1017/jpr.2017.63 Published online by Cambridge University Press

Let U_i denote the number of active and saturated vertices at time i (in other words, U_i is the number of vertices revealed by time i). In particular, by (18), we have

$$U_i = |S_i| + i. \tag{22}$$

Correspondingly, for any vertex u, let $U_{i,r}(u)$ be the number of active and saturated vertices at time i, which are at distance r from u. In particular, for any $i \geq 1$ and any vertex u, it holds that

$$\sum_{r=1}^{N} U_{i,r}(u) = U_i.$$

The number $X_{i,r}$ depends on the number $U_{i-1,r}(v_i)$ of active and saturated vertices at time $i-1$ which are at distance r from v_i, in the following way:

$$X_{i,r}|_{U_{i-1,r}(v_i)} \in \text{binomial}(N_r - U_{i-1,r}(v_i),\, p_r), \tag{23}$$

where we use the notation

$$p_r = \frac{c}{Nr} = p(u,v) \quad \text{if } d(u,v) = r.$$

Remark 2. In (23) and elsewhere, we write a random parameter for a distribution with the usual meaning that the distribution is defined conditionally on a given value of the parameter.

Let us introduce the random variables

$$Z_{i,r} \in \text{binomial}(U_{i-1,r}(v_i),\, p_r), \qquad X^+_{i,r} = X_{i,r} + Z_{i,r} \in \text{binomial}(N_r,\, p_r).$$

Then, we define

$$X^+_i := \sum_{r=1}^{N} X^+_{i,r}.$$

If a random variable ξ stochastically dominates η we denote this by $\eta \preceq \xi$.

It is clear from the above definition that $X_{i,r} \preceq X^+_{i,r}$, and, correspondingly, $X_i \preceq X^+_i$. Therefore,

$$|S_i| \preceq S^+_i := X^+_1 + \cdots + X^+_i - (i-1).$$

Note that the largest connected component has size larger than k if and only if there is a component of size at least k. Then

$$\mathbb{P}\{C(G_N) \geq k\} = \mathbb{P}\{\text{there exists } v\colon |C_v(N)| \geq k\} = \mathbb{P}\left\{\bigcup_{v\in V} \{|C_v(N)| \geq k\right\}.$$

It follows simply by the symmetry of the model that the random variables $|C_v|$, $v \in V_N$, are equally distributed. This allows us to derive, from the last equation, the following bound:

$$\mathbb{P}\{C(G_N) \geq k\} \leq N^2 \mathbb{P}\{|C_v(N)| \geq k\} \tag{24}$$

for any arbitrarily fixed vertex v.

By the exploration algorithm, we find that the probability for a component $C_v(N)$ to be larger or equal to k is equal to the probability of having active vertices in each of the $k-1$ steps of the exploration, hence,

$$\begin{aligned}\mathbb{P}\{|C_v(N)| \geq k\} &= \mathbb{P}\{|S_t| > 0 \text{ for all } t \leq k-1\} \\ &\leq \mathbb{P}\{S^+_t > 0 \text{ for all } t \leq k-1\}\end{aligned}$$

https://doi.org/10.1017/jpr.2017.63 Published online by Cambridge University Press

$$\leq \mathbb{P}\{S_{k-1}^{+} > 0\}$$
$$= \mathbb{P}\left\{\sum_{t=1}^{k-1} X_t^{+} - (k-2) > 0\right\}. \tag{25}$$

We use the coupling method described in [11] for finding stochastic bounds on X_i^{+}. It follows that $X_{i,r}^{+}$ is stochastically bounded from above by a random variable $Y_{i,r} \stackrel{\mathrm{D}}{=} \text{Poisson}(-N_r \log(1-p_r))$, i.e. $X_{i,r}^{+} \preceq Y_{i,r}$. Therefore, we can stochastically bound X_i^{+} by a Poisson random variable as follows:

$$X_i^{+} \preceq \sum_{r=1}^{N} Y_{i,r} \in \text{Poisson}\left(\sum_{r=1}^{N} -N_r \log(1-p_r)\right) = \text{Poisson}(\lambda_N), \tag{26}$$

where

$$\begin{aligned}\lambda_N &= \sum_{r=1}^{N} -N_r \log(1-p_r)\\ &= \sum_{r=1}^{N} N_r(p_r + o(p_r))\\ &= \sum_{r=1}^{\lfloor N/2 \rfloor} 4r(p_r + o(p_r)) + \sum_{r=\lfloor N/2 \rfloor + 1}^{N} 4(N-r)(p_r + o(p_r))\\ &= \lambda - \frac{2c}{N} + o\left(\frac{1}{N}\right).\end{aligned} \tag{27}$$

Let Y_i, $i \geq 1$, be i.i.d. random variables with Poisson(λ_N) distribution. Then we derive, using (25) and (26) with (27), the following upper bound for the probability in (24):

$$\mathbb{P}\{C(G_N) \geq k\} \leq N^2 \mathbb{P}\left\{\sum_{t=1}^{k-1} X_t^{+} > k-2\right\} \leq N^2 \mathbb{P}\left\{\sum_{t=1}^{k-1} Y_t > k-2\right\}. \tag{28}$$

Using Chebyshev's inequality in (28), for any $h > 0$, we have

$$\begin{aligned}\mathbb{P}\{C(G_N) \geq k\} &\leq N^2 \mathbb{P}\left\{\sum_{t=1}^{k-1} Y_t > k-2\right\}\\ &\leq \frac{N^2 \prod_{t=1}^{k-1} \mathbb{E}\mathrm{e}^{hY_t}}{\mathrm{e}^{h(k-2)}}\\ &= N^2 \exp(-h(k-2)) \prod_{t=1}^{k-1} \exp(\lambda_N(\mathrm{e}^h - 1))\\ &= N^2 \exp(-h(k-2)) \exp((k-1)\lambda_N(\mathrm{e}^h - 1)).\end{aligned} \tag{29}$$

The last equation attains its minimum at $h = \log((k-1)/k\lambda)$, where it is equal to

$$N^2 \exp(k(1 - \lambda + \log\lambda) + ko(1)).$$

Therefore, setting $k = (2/(\lambda - 1 - \log\lambda) + \varepsilon)\log N$ in (29), we find that (19) holds.

https://doi.org/10.1017/jpr.2017.63 Published online by Cambridge University Press

Lower bound. To complete the proof of (4), we will prove that, for any $\varepsilon > 0$,

$$\mathbb{P}\left\{\frac{C(G_N)}{\log N} > \frac{2}{1-\lambda-\log\lambda} - \varepsilon\right\} \to 1 \quad \text{as } N \to \infty. \tag{30}$$

Before proceeding to the proof of (30), we derive a useful result, which roughly speaking tells us that removing an arbitrary set of $o(N^2)$ vertices from V_N does not change (asymptotically as $N \to \infty$) the expected degree of a vertex.

Lemma 1. *Let $n_r, r = 1, \ldots, N$, with $0 \le n_r \le N_r$, be an arbitrary sequence such that*

$$\sum_{r=1}^{N} n_r = o(N^2).$$

Then

$$\frac{1}{N}\sum_{r=1}^{N}\frac{n_r}{r} \to 0 \quad \textit{as } N \to \infty.$$

Proof. We prove the lemma by contradiction. Assume there exists a constant $c > 0$ such that, for any $M \in \mathbb{N}$, there exists $N \ge M$ such that

$$\frac{1}{N}\sum_{r=1}^{N}\frac{n_r}{r} \ge c. \tag{31}$$

Let $0 < \delta < \min\{4, c\}$ and define the sets $\mathcal{N}_\delta$ and its complementary $\overline{\mathcal{N}}_\delta$ as follows:

$$\mathcal{N}_\delta = \{r \in \{1, \ldots, N\} : n_r \ge \delta r\}, \qquad \overline{\mathcal{N}}_\delta = \{r \in \{1, \ldots, N\} : n_r < \delta r\}.$$

Noting that from (20) and (21), we have $n_r \le N_r \le 4r$ for any $0 \le r \le N$, from (31) it follows that

$$\begin{aligned} c &\le \frac{1}{N}\sum_{r=1}^{N}\frac{n_r}{r} \\ &= \frac{1}{N}\left(\sum_{r\in\mathcal{N}_\delta}\frac{n_r}{r} + \sum_{r\in\overline{\mathcal{N}}_\delta}\frac{n_r}{r}\right) \\ &\le \frac{1}{N}\left(\sum_{r\in\mathcal{N}_\delta} 4 + \sum_{r\in\overline{\mathcal{N}}_\delta}\delta\right) \\ &= \frac{1}{N}(4|N_\delta| + \delta\overline{\mathcal{N}}_\delta) \\ &= \delta + \frac{4-\delta}{N}|N_\delta|. \end{aligned}$$

In particular, we have

$$|N_\delta| \ge \frac{c-\delta}{4-\delta}N,$$

and, therefore,

$$\sum_{r=1}^{N} n_r \ge \sum_{r\in\mathcal{N}_\delta} n_r \ge \sum_{r\in\mathcal{N}_\delta}\delta r \ge \delta\sum_{r=1}^{|\mathcal{N}_\delta|} r \ge \frac{\delta}{2}|\mathcal{N}_\delta|^2 \ge \frac{\delta}{2}\left(\frac{c-\delta}{4-\delta}\right)^2 N^2,$$

which contradicts the assumptions. □

https://doi.org/10.1017/jpr.2017.63 Published online by Cambridge University Press

Now we can prove (30). We shall follow the construction used already in [17]. For any vertex v, let $V(C_v(N))$ denote here the set of vertices of the component $C_v(N)$. Observe that G_N can be decomposed into pairwise disjoint connected components as follows. Set $\tilde{v}_1 = v$. Then, given $C_{\tilde{v}_1}(N), \dots, C_{\tilde{v}_k}(N)$, for $k \geq 1$ choose a vertex $\tilde{v}_{k+1}$ uniformly in $V_N \setminus \bigcup_{i=1}^{k} V(C_{\tilde{v}_i}(N))$, unless the last set is empty, in which case we stop the algorithm. The graph G_N is thus decomposed into pairwise disjoint connected components $C_{\tilde{v}_1}(N), \dots, C_{\tilde{v}_M}(N)$, where $M = M(N)$ is a bounded random variable, $1 \leq M \leq N^2$, denoting the number of disjoint components in G_N.

Fix $\varepsilon > 0$ arbitrarily and denote $K_N = (2/(\lambda - 1 - \log\lambda) + \varepsilon)\log N$. Then we define the event

$$E_N = \{C(G_N) \leq K_N\}.$$

Recall that, from (19), it follows that

$$\mathbb{P}\{\bar{E}_N\} \to 0 \quad \text{as } N \to \infty.$$

This yields, for any $k \geq 1$,

$$\begin{aligned}\mathbb{P}\{C(G_N) \leq k\} &= \mathbb{P}\{|C_{\tilde{v}_1}(N)| \leq k, \dots, |C_{\tilde{v}_M}(N)| \leq k\} \\ &\leq \mathbb{P}\{|C_{\tilde{v}_1}(N)| \leq k, \dots, |C_{\tilde{v}_M}(N)| \leq k \mid E_N\} + o(1). \end{aligned} \tag{32}$$

Note that, since conditionally on E_N the largest connected component is smaller than K_N, it follows that $MK_N \geq N^2$. Hence, for any $m_N \leq N^2/K_N \leq M$, the following bound holds for the probability in (32):

$$\begin{aligned}&\mathbb{P}\{|C_{\tilde{v}_1}(N)| \leq k, \dots, |C_{\tilde{v}_M}(N)| \leq k \mid E_N\} \\ &\quad\leq \prod_{i=1}^{m_N} \mathbb{P}\{|C_{\tilde{v}_i}(N)| \leq k \mid |C_{\tilde{v}_1}(N)| \leq k, \dots, |C_{\tilde{v}_{i-1}}(N)| \leq k, E_N\}. \end{aligned} \tag{33}$$

Let $\mathcal{V}_0$ be an arbitrary set of $m_N K_N$ nodes, u be a vertex in $V_N \setminus \mathcal{V}_0$, and let $\tilde{C}_u = \tilde{C}_u(\mathcal{V}_0)$ denote the connected component containing u constructed precisely as the original $C_v(N)$ but on the vertex set $V_N \setminus \mathcal{V}_0$.

Then each factor in (33) can be uniformly bounded as follows:

$$\mathbb{P}\{|C_{\tilde{v}_i}(N)| \leq k \mid |C_{\tilde{v}_1}(N)| \leq k, \dots, |C_{\tilde{v}_{i-1}}(N)| \leq k, E_N\} \leq \max_{\substack{\mathcal{V}_0 \subseteq V_N : |\mathcal{V}_0| = m_N K_N \\ u \in V_N \setminus \mathcal{V}_0}} \mathbb{P}\{|\tilde{C}_u| \leq k\},$$

where we simply used the fact that on a smaller set of vertices, the components are smaller as well. Therefore, from (33), it follows that

$$\mathbb{P}\{C(G_N) \leq k\} \leq \Big(\max_{\mathcal{V}_0 :\, |\mathcal{V}_0| = m_N K_N,\, u \in V_N \setminus \mathcal{V}_0} \mathbb{P}\{|\tilde{C}_u| \leq k\}\Big)^{m_N}. \tag{34}$$

In the following, fix the set $\mathcal{V}_0 \subset V_N$ arbitrarily but so that

$$|\mathcal{V}_0| = m_N K_N = o(N^2).$$

Fix a vertex $u \notin \mathcal{V}_0$ arbitrarily, and construct the component $\tilde{C}_u$ on the vertex set $V_N \setminus \mathcal{V}_0$ as described in the exploration algorithm. Let us denote here $u_1, u_2, \dots$, the sequence of saturated vertices (which corresponds to the sequence $v_1, v_2, \dots$, in the original exploration algorithm).

Define $n_r^0(u)$ to be the number of nodes in $\mathcal{V}_0$ which are at distance r from u, so that $0 \leq n_r^0(u) \leq N_r$ and $\sum_{r=1}^{N} n_r^0(u) = |\mathcal{V}_0|$ for any u.

https://doi.org/10.1017/jpr.2017.63 Published online by Cambridge University Press

Analogous to the notion used previously, let $\tilde{U}_i$ here denote the number of active and saturated vertices at step i in this new exploration process on $V_N \setminus \mathcal{V}_0$ (see (22)), and $\tilde{U}_{i,r}(w)$ be the number of those vertices at distance r from the vertex w. Let also $n^0_{i,r} = n^0_r(u_i)$ denote the number of the vertices in $\mathcal{V}_0$ which are at distance r from the ith saturated vertex u_i. By this definition, and our assumption on $|\mathcal{V}_0| = o(N^2)$, we have

$$\sum_{r=1}^{N} n^0_{i,r} = |\mathcal{V}_0| = m_N K_N = o(N^2) \quad \text{for any } i. \tag{35}$$

Hence, the number of vertices activated at step i at distance r from the ith explored vertex, which we denote $\tilde{X}_{i,r}$, has the following distribution:

$$\tilde{X}_{i,r} \in \text{binomial}(N_r - n^0_{i,r} - \tilde{U}_{i-1,r}(u_i), p_r),$$

and the total number of vertices activated at the ith step is given by

$$\tilde{X}_i = \sum_{r=1}^{N} \tilde{X}_{i,r}.$$

Using these definitions, we derive, for any $k \geq 1$,

$$\mathbb{P}\{|\tilde{C}_u| > k\} \geq \mathbb{P}\{\tilde{X}_1 + \tilde{X}_2 + \cdots + \tilde{X}_t > t - 1 \text{ for all } t \leq k - 1\}. \tag{36}$$

To approximate the distribution of $\tilde{X}_i$ in the last equation, let us recall the following result on the Poisson approximation.

Lemma 2. (See, e.g. [21].) *On a rich enough probability space, we can define a random vector* (X, Y) *so that* $X \overset{\text{D}}{=} \text{binomial}(n, \lambda/n)$, $Y \overset{\text{D}}{=} \text{Poisson}(\lambda)$, *and, moreover,*

$$\mathbb{P}(X \neq Y) \leq \frac{\lambda^2}{n}.$$

Given the numbers $0 \leq k_{i,r} \leq N_r - n^0_{i,r}$, $r = 1, \ldots, N$, $i = 1, \ldots, k-1$, such that

$$\sum_{r=1}^{N} k_{i,r} \leq k, \tag{37}$$

let us define couplings $(\tilde{X}_{i,r}, \tilde{Z}_{i,r})$ with the Poisson random variables

$$\tilde{Z}_{i,r} \in \text{Poisson}((N_r - n^0_{i,r} - k_{i,r})p_r),$$

which satisfy the conditions in Lemma 2.

Then

$$\tilde{Z}_i = \sum_{r=1}^{N} \tilde{Z}_{i,r} \in \text{Poisson}(\lambda_{i,N}), \tag{38}$$

where

$$\lambda_{i,N} = \sum_{r=1}^{N} (N_r - n^0_{i,r} - k_{i,r})p_r. \tag{39}$$

To simplify the notation, define the event

$$F_N = \{\tilde{U}_{i,r} = k_{i,r}, \sum_{r=1}^{N} k_{i,r} \leq k \text{ for all } i \leq k-1,\ r = 1, \ldots, N\}$$

https://doi.org/10.1017/jpr.2017.63 Published online by Cambridge University Press

and consider

$$\begin{aligned}\mathbb{P}(\tilde{Z}_i > k \mid F_N) &= \mathbb{P}(\tilde{Z}_i > k, \tilde{Z}_i = \tilde{X}_i \mid F_N) + \mathbb{P}(\tilde{Z}_i > k, \tilde{Z}_i \neq \tilde{X}_i \mid F_N) \\ &\leq \mathbb{P}(\tilde{X}_i > k \mid F_N) + \mathbb{P}(\tilde{Z}_i \neq \tilde{X}_i \mid F_N).\end{aligned}$$

Note that

$$\begin{aligned}\mathbb{P}\{\tilde{X}_i \neq \tilde{Z}_i \mid F_N\} &= \mathbb{P}\left\{\sum_{r=1}^{N} \tilde{X}_{i,r} \neq \sum_{r=1}^{N} \tilde{Z}_{i,r} \;\middle|\; F_N\right\} \\ &\leq \mathbb{P}\left\{\bigcup_{r=1}^{N} \{\tilde{X}_{i,r} \neq \tilde{Z}_{i,r}\} \;\middle|\; F_N\right\} \\ &\leq \sum_{r=1}^{N} \mathbb{P}\{\tilde{X}_{i,r} \neq \tilde{Z}_{i,r} \mid F_N\}.\end{aligned}$$

By Lemma 2, we have

$$\mathbb{P}\{\tilde{X}_{i,r} \neq \tilde{Z}_{i,r} \mid F_N\} \leq p_r^2(N_r - n_{i,r}^0 - k_{i,r}),$$

which yields

$$\begin{aligned}\mathbb{P}\{\tilde{X}_i \neq \tilde{Z}_i \mid F_N\} &\leq \sum_{r=1}^{N} p_r^2(N_r - n_{i,r}^0 - k_{i,r}) \\ &= \frac{c^2}{N^2} \sum_{r=1}^{N} \frac{1}{r^2}(N_r - n_{i,r}^0 - k_{i,r}) \\ &= O\left(\frac{\log N}{N^2}\right), \qquad (40)\end{aligned}$$

uniformly in i.

Next we consider

$$\begin{aligned}&\mathbb{P}\{\tilde{Z}_1 + \cdots + \tilde{Z}_t > t - 1 \text{ for all } t \leq k - 1\} \\ &\quad \leq \mathbb{P}\{\tilde{X}_1 + \cdots + \tilde{X}_t > t - 1 \text{ for all } t \leq k - 1 \mid F_N\} + \sum_{s=1}^{k-1} \mathbb{P}\{\tilde{X}_s \neq \tilde{Z}_s \mid F_N\}. \qquad (41)\end{aligned}$$

Note that, by (40),

$$\varepsilon_k(N) := \sum_{s=1}^{k-1} \mathbb{P}\{\tilde{X}_s \neq \tilde{Z}_s \mid F_N\} = O\left(\frac{k \log N}{N^2}\right). \qquad (42)$$

Therefore, (41) yields

$$\begin{aligned}&\mathbb{P}\{\tilde{X}_1 + \cdots + \tilde{X}_t > t - 1 \text{ for all } t \leq k - 1 \mid F_N\} \\ &\quad \geq \mathbb{P}\{\tilde{Z}_1 + \cdots + \tilde{Z}_t > t - 1 \text{ for all } t \leq k - 1\} - \varepsilon_k(N). \qquad (43)\end{aligned}$$

We shall construct i.i.d. random variables $\tilde{Z}_i^-$, $1 \leq i \leq k$, which are a.s. smaller than $\tilde{Z}_i$, $1 \leq i \leq k$, correspondingly.

https://doi.org/10.1017/jpr.2017.63 Published online by Cambridge University Press

First, using assumption (35) and Lemma 1, we derive

$$\sum_r n_{i,r}^0 p_r = \frac{c}{N} \sum_r \frac{n_{i,r}^0}{r} = o(1). \tag{44}$$

From now on, we shall assume that

$$k = a \log N \quad \text{for some positive } a. \tag{45}$$

Under this assumption, we have

$$\sum_r k_{i,r} p_r = \frac{c}{N} \sum_{r=1}^N \frac{k_{i,r}}{r} \leq \frac{c}{N} \sum_r k_{i,r} = \frac{kc}{N} = \frac{ac \log N}{N} = o(1). \tag{46}$$

Hence, from (44) and (46), we obtain the following bound for $\lambda_{i,N}$ defined in (39):

$$\lambda_{i,N} = \sum_{r=1}^N (N_r - n_{i,r}^0 - k_{i,r}) p_r \geq \sum_{r=1}^N N_r p_r + o_i(1), \tag{47}$$

where $o_i(1)$ might depend on i. Note that, by (27),

$$\sum_{r=1}^N N_r p_r = \lambda + o(1). \tag{48}$$

Hence, for any (constant)

$$\lambda' < \lambda, \tag{49}$$

(47) together with (48) yields the following uniform in $i \leq k$ bound:

$$\lambda_{i,N} > \lambda'. \tag{50}$$

Recall that $\tilde{Z}_i \in \text{Poisson}(\lambda_{i,N})$ by (38). Therefore, (50) allows us to construct independent $\tilde{Z}_i^- \in \text{Poisson}(\lambda')$, $1 \leq i \leq k$, such that

$$\tilde{Z}_i^- \leq \tilde{Z}_i \quad \text{a.s. for each } i.$$

Now we can derive the following bound:

$$\begin{aligned} \mathbb{P}(\tilde{Z}_1 + \ldots, \tilde{Z}_t > t-1,\ t = 1, \ldots, k-1) &\geq \mathbb{P}(\tilde{Z}_1^- + \cdots + \tilde{Z}_t^- > t-1,\ t = 1, \ldots, k-1) \\ &= \mathbb{P}\{\mathcal{T} \geq k\}, \end{aligned}$$

where $\mathcal{T}$ denotes the total progeny of a branching process with offspring distribution $\text{Poisson}(\lambda')$. Substituting this bound into (43), we obtain

$$\mathbb{P}\{\tilde{X}_1 + \cdots + \tilde{X}_t > t-1 \text{ for all } t \leq k-1 \mid F_N\} \geq \mathbb{P}\{\mathcal{T} \geq k\} - \varepsilon_k(N),$$

where the right-hand side is uniform in F_N (here we still assume conditions (37) and (45)). This yields

$$\mathbb{P}\{\tilde{X}_1 + \cdots + \tilde{X}_t > t-1 \text{ for all } t \leq k-1\} \geq \mathbb{P}\{\mathcal{T} \geq k\} - \varepsilon_k(N),$$

https://doi.org/10.1017/jpr.2017.63 Published online by Cambridge University Press

and, therefore, by (36),

$$\mathbb{P}\{|\tilde{C}_u| \le k\} \le 1 - \mathbb{P}\{\mathcal{T} \ge k\} + \varepsilon_k(N). \tag{51}$$

Using a well-known formula for the distribution of the progeny of a branching process (see, e.g. [13]), we compute

$$\mathbb{P}\{\mathcal{T} \ge k\} = \sum_{j=k}^{\infty} \frac{(\lambda' j)^{j-1}}{j!} \mathrm{e}^{\lambda' j} \ge \frac{(\lambda' k)^{k-1}}{k!} \mathrm{e}^{\lambda' k},$$

which, together with the Stirling formula, yields

$$\mathbb{P}\{\mathcal{T} \ge k\} \ge \frac{1}{\sqrt{2\pi}\lambda'} \frac{1}{k^{3/2}} \mathrm{e}^{-\alpha k} \left(1 + O\left(\frac{1}{k}\right)\right), \tag{52}$$

where

$$\alpha = \lambda' - 1 - \log \lambda'. \tag{53}$$

Substituting (52) into (51), we obtain, using (42) for $k = a \log N$,

$$\mathbb{P}\{|\tilde{C}_u| \le a \log N\} \le 1 - \frac{1}{A_N}(1 + o(1)) + O\left(\left(\frac{\log N}{N}\right)^2\right), \tag{54}$$

where

$$A_N = \sqrt{2\pi}\lambda'(a \log N)^{3/2} N^{a\alpha}.$$

Choose now arbitrarily a constant

$$a < \frac{2}{\alpha}. \tag{55}$$

Then (54) yields

$$\mathbb{P}\{|\tilde{C}_u| \le a \log N\} \le 1 - \frac{1}{A_N}(1 + o(1)). \tag{56}$$

Observe that the value on the right-hand side of the above equation is uniform in the choice of the set $\mathcal{V}_0$ and vertex u. Therefore, we can use bound (56) in (34) to obtain

$$\mathbb{P}\{C(G_N) \le a \log N\} \le \left(1 - \frac{1}{A_N}(1 + o(1))\right)^{m_N}. \tag{57}$$

Finally, we choose

$$m_N = A_N \log N \gg A_N,$$

which by (55) also satisfies condition (35), i.e.

$$m_N K_N = o(N^2),$$

where

$$K_N = \left(\frac{2}{\lambda - 1 - \log \lambda} + \varepsilon\right) \log N.$$

With this choice of m_N, bound (57) implies that

$$\mathbb{P}\{C(G_N) \le a \log N\} = o(1) \tag{58}$$

for any fixed constant (see (53) and (55))

$$a < \frac{2}{\alpha} = \frac{2}{\lambda' - 1 - \log \lambda'}.$$

https://doi.org/10.1017/jpr.2017.63 Published online by Cambridge University Press

By (49), here we can choose any $\lambda' < \lambda$; therefore, it follows that (58) holds for any

$$a < \frac{2}{\lambda - 1 - \log \lambda}.$$

This proves (30), and, therefore, part (i) of Theorem 1 is proved.

This completes the proof of Theorem 1, since part (ii) follows by Theorem 2. □

4.4. Outline of the proof of Theorem 3

A proof of Theorem 3 can be obtained by following the same strategy as in the proof of Theorem 1(i), in combination with the proof of the corresponding result for the rank-1 model (8) given in [17]. Therefore, we omit the details here.

Acknowledgements

The work has been partially supported by the Swedish Research Council and by the Crafoord Foundation. We thank the anonymous referees for the constructive remarks.

References

[1] Aizenman, M. and Newman, C. M. (1986). Discontinuity of the percolation density in one dimensional $1/|x-y|^2$ percolation models. *Commun. Math. Phys.* **107,** 611–647.
[2] Alon, N. and Spencer, J. H. (1992). *The Probabilistic Method. With an Appendix by Paul Erdős.* John Wiley, New York.
[3] Avin, C. (2008). Distance graphs: from random geometric graphs to Bernoulli graphs and between. In *DialM-POMC '08*, 71–77.
[4] Bhamidi, S., van der Hofstad, R. and van Leeuwaarden, J. S. H. (2010). Scaling limits for critical inhomogeneous random graphs with finite third moments. *Electron. J. Prob.* **15,** 1682–1702.
[5] Bhamidi, S., van der Hofstad, R. and van Leeuwaarden, J. S. H. (2012). Novel scaling limits for critical inhomogeneous random graphs. *Ann. Prob.* **40,** 2299–2361.
[6] Bollobás, B., Janson, S. and Riordan, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31,** 3–122.
[7] Bollobás, B., Janson, S. and Riordan, O. (2007). Spread-out percolation in $\mathbb{R}^d$. *Random Structures Algorithms* **31,** 239–246.
[8] Deijfen, M., van der Hofstad, R. and Hooghiemstra, G. (2013). Scale-free percolation. *Ann. Inst. H. Poincaré Prob. Statist.* **49,** 817–838.
[9] Janson, S. (2008). The largest component in a subcritical random graph with a power law degree distribution. *Ann. Appl. Prob.* **18,** 1651–1668.
[10] Janson, S., Kozma, R., Ruszinkó, M. and Sokolov, Y. (2015). Bootstrap percolation on a random graph coupled with a lattice. Preprint. Available at https://arxiv.org/abs/1507.07997v2.
[11] Lindvall, T. (1992). *Lectures on the Coupling Method*. John Wiley, New York.
[12] Markram, H. *et al.* (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell* **163,** 456–492.
[13] Otter, R. (1949). The multiplicative process. *Ann. Math. Statist.* **20,** 206–224.
[14] Penrose, M. D. (1993). On the spread-out limit for bond and continuum percolation. *Ann. Appl. Prob.* **3,** 253–276.
[15] Penrose, M. (2003). *Random Geometric Graphs*. Oxford University Press.
[16] Reimann, M. W. *et al.* (2017). Cliques of neurons bound into cavities provide a missing link between structure and function. *Front. Comput. Neurosci.* **11,** 48pp.
[17] Turova, T. S. (2011). The largest component in subcritical inhomogeneous random graphs. *Combin. Prob. Comput.* **20,** 131–154.
[18] Turova, T. S. (2012). The emergence of connectivity in neuronal networks: from bootstrap percolation to auto-associative memory. *Brain Res.* **1434,** 277–284.
[19] Turova, T. S. (2013). Asymptotics for the size of the largest component scaled to 'log n' in inhomogeneous random graphs. *Ark. Mat.* **51,** 371–403.
[20] Turova, T. S. (2013). Diffusion approximation for the components in critical inhomogeneous random graphs of rank-1. *Random Structures Algorithms* **43,** 486–539.
[21] van der Hofstad, R. (2016). *Random Graphs and Complex Networks,* Vol. 1. Cambridge University Press.