Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2011, Vol. 47, No. 4, 929–968
DOI: 10.1214/10-AIHP407
© Association des Publications de l'Institut Henri Poincaré, 2011

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Giant vacant component left by a random walk in a random d-regular graph[1]

Jiří Černý[a], Augusto Teixeira[a] and David Windisch[b]

[a]*Department of Mathematics, ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland. E-mail: cerny@math.ethz.ch; teixeira@math.ethz.ch*
[b]*Faculty of Mathematics and Computer Science, The Weizmann Institute of Science, POB 26, Rehovot 76100, Israel. E-mail: david.windisch@weizmann.ac.il*

Received 1 December 2009; revised 3 September 2010; accepted 11 December 2010

Abstract. We study the trajectory of a simple random walk on a d-regular graph with $d \geq 3$ and locally tree-like structure as the number n of vertices grows. Examples of such graphs include random d-regular graphs and large girth expanders. For these graphs, we investigate percolative properties of the set of vertices not visited by the walk until time un, where $u > 0$ is a fixed positive parameter. We show that this so-called *vacant* set exhibits a phase transition in u in the following sense: there exists an explicitly computable threshold $u_\star \in (0, \infty)$ such that, with high probability as n grows, if $u < u_\star$, then the largest component of the vacant set has a volume of order n, and if $u > u_\star$, then it has a volume of order $\log n$. The critical value $u_\star$ coincides with the critical intensity of a random interlacement process on a d-regular tree. We also show that the random interlacements model describes the structure of the vacant set in local neighbourhoods.

Résumé. Nous étudions la trajectoire d'une marche aléatoire simple sur un graphe d-régulier avec $d \geq 3$ dont la structure ressemble localement à un arbre, quand le nombre de sommets n du graphe croît. Des exemples de tels graphes comprennent des graphes aléatoires d-réguliers et des 'expanseur de grande maille'. Pour ces graphes, nous étudions les propriétés de percolation de l'ensemble des sommets non visités par la marche jusqu'au moment un, où $u > 0$ est un paramètre positif fixé. Nous montrons que cet *ensemble vacant* subit une transition de phase en u dans le sens suivant : il existe un seuil $u_\star \in (0, \infty)$ explicitement calculable tel que, avec une forte probabilité quand n croît, si $u < u_\star$, la plus grande composante de l'ensemble vacant a un volume d'ordre n, et si $u > u_\star$, elle a un volume d'ordre $\log n$. La valeur critique $u_\star$ coïncide avec l'intensité critique des entrelacs aléatoires sur un arbre d-régulier. Nous montrons aussi que les entrelacs aléatoires décrivent bien la structure de l'ensemble vacant dans des voisinages locaux.

MSC: 60G50; 05C80; 82B41

Keywords: Random walk; Vacant set; Regular graph; Expanders; Random interlacement; Phase transition

1. Introduction and main results

In this work we consider the simple random walk on a graph G chosen among a certain class of finite regular graphs including, for example, typical realizations of random regular graphs, or expanders with large girth. The main object of our study is the complement of the trajectory of the random walk stopped at a time $u|G|$, for $u > 0$, the so-called vacant set, and its percolative properties.

We show that the vacant set undergoes the following phase transition in u: as long as $u < u_\star$, the vacant set has a unique component with volume of order $|G|$, whereas if $u > u_\star$, the largest component of the vacant set only has

[1]The research of A. Teixeira and D. Windisch was partially supported by Swiss National Science Foundation Grant PDFM22-120708/1.

a volume of order $\log|G|$, with high probability as the size of G diverges. More importantly, we show that the above phase transition corresponds to the phase transition in a random interlacement model on a regular tree. In particular, the critical value $u_\star$ is the same for both models.

The random interlacement (on $\mathbb{Z}^d$, $d \geq 3$) was recently introduced by Sznitman [25] to provide a model describing the microscopic structure of the bulk when considering the large N asymptotics of the disconnection time of the discrete cylinder $(\mathbb{Z}/N\mathbb{Z})^{d-1} \times \mathbb{Z}$ [9], or percolative properties of the vacant set left by the simple random walk on the torus $(\mathbb{Z}/N\mathbb{Z})^d$ [6]. Later, in [23,28], it was proved that the random interlacements indeed *locally* describe this microscopic structure. In [20,25] it was shown that the random interlacement undergoes a phase transition for a non-trivial value $u_\star(\mathbb{Z}^d)$ of the parameter u driving its intensity. The best bounds on the disconnection time known from [22] and [24] involve parameters derived from random interlacements and the upcoming work [27] connects distinct regimes for random interlacements with distinct regimes for the vacant set left by random walk on the torus. It can currently not be proved that the critical value $u_\star(\mathbb{Z}^d)$ for random interlacements is itself connected with a critical value for the vacant set on the torus or for the disconnection time.

We provide such a connection in our model. This is possible for the following reasons: For the considered graphs G, large neighbourhoods of typical vertices of G are isomorphic to a ball in a regular tree, and, as we will show, the corresponding local microscopic model is the random interlacement on such a tree. Connected components of this interlacement model admit a particularly simple description in terms of a branching process, and its critical value $u_\star$ is explicitly computable (see (1.2)), giving us a good local control of configurations of the vacant set on G. Good expansion properties of G then allow us to extend the local control to a global one.

We now come to the precise statements of our results. We consider a sequence of finite connected graphs $G_k = (V_k, \mathcal{E}_k)$ such that the number n_k of vertices in V_k tends to infinity as $k \to \infty$. We are principally interested in the case where G_k is a sequence of d-regular random graphs, $d \geq 3$, or of d-regular expanders with large girth (such as, for example, Lubotzky–Phillips–Sarnak graphs [15]). As we shall show below, these two classes of graphs satisfy the following assumptions, which are the only assumptions we need in order to prove our main theorems. We assume that for some $d \geq 3$, $\alpha_1 \in (0, 1)$, and all k,

$$G_k \text{ is } d\text{-regular, that is all its vertices have degree } d\text{, and} \tag{A0}$$

$$\text{for any } x \in V_k\text{, there is at most one cycle contained in the ball with radius } \alpha_1 \log_{d-1} n_k \text{ centered at } x. \tag{A1}$$

We also assume that the spectral gap λ_{G_k} of G_k (we recall the definition in (2.7) below) is uniformly bounded from below by a constant $\alpha_2 > 0$, that is

$$\lambda_{G_k} > \alpha_2 > 0 \quad \text{for all } k \geq 1. \tag{A2}$$

Under (A0), this final assumption is equivalent to assuming that G_k are expanders, see (2.11). Note that in general (A1) does not imply (A2), see Remark 1.5.

We consider a continuous-time random walk on G_k. More precisely, we write P for the canonical law on the space $D([0, \infty), V_k)$ of cadlag functions from $[0, \infty)$ to V_k of the continuous-time simple random walk on G_k with i.i.d. mean-one exponentially distributed waiting times and uniformly distributed starting point. We use $(X_t)_{t \geq 0}$ to denote the canonical coordinate process. For a fixed parameter $u \geq 0$ not depending on k, we define the *vacant set* as the set of all vertices not visited by the random walk until time un_k:

$$\mathcal{V}_k^u = \{x \in V_k\colon\ x \neq X_t \text{ for all } 0 \leq t \leq un_k\}. \tag{1.1}$$

We use $\mathcal{C}_{\max}^u \subset V$ to denote the largest connected component of $\mathcal{V}_k^u$.

The following theorems are the main results of the present paper. The critical parameter $u_\star$ in the statements coincides with the critical parameter for random interlacements on the infinite d-regular tree $\mathbb{T}_d$, which, according to [26], equals

$$u_\star = \frac{d(d-1)\ln(d-1)}{(d-2)^2}. \tag{1.2}$$

Theorem 1.1 (Subcritical phase). *Assume* (A0)–(A2), *and fix* $u > u_\star$. *Then for every* $\sigma > 0$ *there exist constants* $K(d,\sigma,u,\alpha_1,\alpha_2), C(d,\sigma,u,\alpha_1,\alpha_2) < \infty$, *such that*

$$P\big[\big|\mathcal{C}^u_{\max}\big| \geq K \ln n_k\big] \leq C n_k^{-\sigma} \quad \textit{for all } k \geq 1. \tag{1.3}$$

Theorem 1.2 (Supercritical phase). *Assume* (A0)–(A2), *and fix* $u < u_\star$. *Then for every* $\sigma > 0$ *there exist constants* $\rho(d,\sigma,u,\alpha_1,\alpha_2) \in (0,1)$ *and* $C(d,\sigma,u,\alpha_1,\alpha_2) < \infty$, *such that*

$$P\big[\big|\mathcal{C}^u_{\max}\big| \geq \rho n_k\big] \geq 1 - C n_k^{-\sigma} \quad \textit{for all } k \geq 1. \tag{1.4}$$

For the statement on the uniqueness of the giant component we denote the second largest component of $\mathcal{V}^u_k$ by $\mathcal{C}^u_{\mathrm{sec}}$.

Theorem 1.3 (Supercritical phase-uniqueness). *Assume* (A0)–(A2), *and fix* $u < u_\star$. *Then for every* $\kappa > 0$,

$$\lim_{k\to\infty} P\big[\big|\mathcal{C}^u_{\mathrm{sec}}\big| \geq \kappa n_k\big] = 0. \tag{1.5}$$

From the last theorem it follows that there exists a function f satisfying $f(n) = \mathrm{o}(n)$ and $P[|\mathcal{C}^u_{\mathrm{sec}}| \leq f(n_k)] \to 1$. More information on the asymptotics of $f(n)$ could be obtained from our techniques. However, they are not sufficient to prove $f = \mathrm{O}(\log n)$, which is the conjectured size of $\mathcal{C}^u_{\mathrm{sec}}$, based on the behaviour of Bernoulli percolation.

Let us now comment on related results. The size of the vacant components left by a random walk on a finite graph has so far only been studied by Benjamini and Sznitman in [6] for G_k given by a d-dimensional integer torus with large side length k and sufficiently large dimension d. In this case, the authors prove that the vacant set has a suitably defined unique giant component occupying a non-degenerate fraction of the total volume with overwhelming probability, provided $u > 0$ is chosen sufficiently small. Their work does not prove anything, however, for the large u regime, let alone any results on a phase transition in u. Our results are the first ones to establish such a phase transition for a random walk on a finite graph. Moreover, our results provide some indication that a phase transition occurs for random walk on the torus as well, and that the critical parameter $u_\star(\mathbb{Z}^d)$ for random interlacements on $\mathbb{Z}^d$ should play a key role.

A similar phase transition was proved for Bernoulli percolation on various graphs: first by Erdős and Rényi [11] on the complete graph, and more recently on large-girth expanders in [3], as well as on many other graphs satisfying a so-called triangle condition [7]. For our results the paper [3] is the most relevant, some of our proofs build on techniques introduced there. A very precise description of the Bernoulli percolation on random regular graphs was recently obtained in [17,18].

Let us now comment on the proofs of our results. For most of the arguments, we do not work with the law P of the random walk, but with a different measure Q on $D([0,\infty), V_k)$. The trajectory of the canonical process X under Q is constructed from an i.i.d. sequence $(Y^i)_{i\in\mathbb{N}}$, of uniformly-started random walk trajectories of length $L = n_k^\gamma$, for $\gamma < 1$, called *segments*. To create a nearest-neighbour path, the endpoint of segment Y^i and the starting point of segment Y^{i+1} are connected using a *bridge* Z^i, $i \in \mathbb{N}$, which is a random walk bridge of length $\ell = \log^2 n_k$. Since ℓ is much larger than the mixing time of the random walk on G_k, Q provides a very good approximation of P, see Lemma 4.1.

The set $\bar{\mathcal{V}}^u_k = V_k \setminus \bigcup_{i<\lfloor un_k/(L+\ell)\rfloor} \operatorname{Ran} Y^i$, the so-called *vacant set left by segments*, plays a particular role in our proofs. It is a complement of 'a cloud of independent random walk trajectories,' similar to the vacant set of a random interlacement. Observe that $\bar{\mathcal{V}}^u_k$ is an enlargement of $\mathcal{V}^u_k$.

To prove Theorem 1.1, we analyse a breadth-first search algorithm exploring one component of the set $\bar{\mathcal{V}}^u_k$. We show that this algorithm is likely to terminate in no more than $K \ln n_k$ steps. To prove this, we need to control the probability that a (not yet explored) vertex y is found to be vacant at a particular step of the algorithm. The main difficulty is that, unlike in Bernoulli site percolation models, this event is not independent of the past of the algorithm. We will derive an estimate of the form (see Proposition 3.6)

$$\mathsf{P}\big[y \notin \operatorname{Ran} Y^i \,|\, A \cap \operatorname{Ran} Y^i = \varnothing\big] \sim f(d,u)^{(L+\ell)/(un)}, \tag{1.6}$$

where A will be the part of $\bar{\mathcal{V}}^u_k$ already explored by the algorithm. The explicitly computable quantity $f(d,u)$ appears in the study of random interlacement on the infinite tree $\mathbb{T}_d$ at level u [26], and it equals the probability that a given vertex z (different from the root of the tree) is vacant, given its parent in the tree is vacant.

The estimate (1.6) will imply that the probability of y being vacant given the past of the algorithm is well approximated by $f(d,u)$. Since for $u > u_\star$ we have $f(d,u) < 1/(d-1)$, the considered breadth-first search algorithm can be controlled by a sub-critical branching process, yielding Theorem 1.1.

There is an additional difficulty coming from the fact that the estimate (1.6) holds only under suitable restrictions on the set A and the vertex y (see (3.28)). These restrictions are however always satisfied for a large majority of the steps of the algorithm, as we will show in Proposition 5.4.

We now comment on the proof of Theorem 1.2. This proof consists of the following two steps: first we show that for some slightly larger parameter $u_k \in (u, u_\star)$, there are many components of $\mathcal{V}_k^{u_k}$ having volume at least n_k^δ, for some $\delta > 0$. Then, we use a sprinkling technique, based on the following heuristic idea: we reduce u_k to u and prove that with high probability, the mentioned components merge into a cluster of size at least ρn_k, cf. [3].

For the first step, we use the fact that $\bar{\mathcal{V}}_k^{u_k}$ can be locally compared with the vacant set of random interlacements on a d-regular tree. This is proved in Proposition 6.3, which again uses an approximation of type (1.6), see (6.38). Since $u_k < u_\star$, the random interlacement at level u_k is super-critical, yielding the existence of components of volume n_k^δ in G_k. Lemma 6.9 then implies that going from $\bar{\mathcal{V}}_k^{u_k}$ to $\mathcal{V}_k^{u_k}$ (by inserting the bridges Z^i) does not destroy these components.

Regarding the second step, it is by no means obvious how to perform a sprinkling as mentioned above. Indeed, a simple deletion of the last part $X_{[un_k, u_k n_k]}$ of the trajectory would require us to deal with the distribution of the set $X_{[un_k, u_k n_k]}$ given $\mathcal{V}_k^{u_k}$, which seems difficult. Instead, we perform the sprinkling in the manner natural for random interlacements (cf. [21]): we remove some segments Y^i independently at random.

The deletion of segments, however, disconnects the trajectory of the process. We bypass this problem by adding extra bridges before the sprinkling (cf. (6.52) and Lemma 6.9 again), so that even after the deletion of some segments, we can extract a nearest-neighbour trajectory of length at least un, with high probability.

We then use the expansion properties (cf. (2.11)) of our graph to show that the sprinkling construction merges some of the clusters of size n_k^δ into a giant component of size at least ρn_k.

The proof of the uniqueness, that is of Theorem 1.3, again combines sprinkling with the local comparison with random interlacements. Using this comparison and the branching process approximation of the random interlacement on the tree, we will show that at level u_k there are, with a high probability, only $\mathrm{o}(n_k)$ vertices contained in vacant clusters of size between $\ln^2 n_k$ and n_k^c, for some $c \in (0,1)$, see Lemma 7.1 for the exact formulation. This statement is a weaker version of the so-called 'absence of components of intermediate size' which is usually proved for Bernoulli percolation.

This will allow us to show that any component of $\mathcal{V}_k^u$ of size at least κn_k, should contain at least $\kappa n_k/2$ vertices x being in vacant components of size at least n_k^c at level u_k. The sprinkling then shows that any two groups of size $\kappa n_k/2$ of such vertices are connected in $\mathcal{V}_k^u$, excluding two giant components with a high probability.

We close this introduction with two remarks concerning our assumptions.

Remark 1.4. *The assumptions* (A0)–(A2) *are designed in order to include two classes of d-regular graphs*: *expanders with girth larger than $c \log |V_k|$, and typical realizations of a random d-regular graph. In the random d-regular graph case these assumptions also help us separate the randomness of the graph from the randomness of the walk.*

The fact that the typical realization of the random d-regular graph satisfies assumption (A1) *follows from Lemma* 2.1 *of* [14], *where they show it for $\alpha_1 = \frac{1}{5}$. To see that* (A2) *holds one can use the estimate on the second eigenvalue of the adjacency matrix A of the random d-regular graph of Friedman* [13] (*or older results*, *e.g.* [8,12], *which however only provide estimates for d even and not too small*). *Indeed*, *in* [13] *it is shown that this second eigenvalue is $2\sqrt{d-1} + \mathrm{o}(1)$, with a high probability. The largest eigenvalue of this matrix is d. This implies* (A2), *since the generator of the random walk is given by $\frac{A}{d} - \mathrm{Id}$.*

Remark 1.5. *The assumption* (A1) *does not imply* (A2). *This can be seen easily by considering two copies G, G' of a large girth expander with n vertices*, *choosing two edges*, $e = \{x, y\}$ *of G and $e' = \{x', y'\}$ of G'*, *erasing e, e' and joining G, G' with two new edges $\{x, x'\}$*, *and $\{y, y'\}$. The new graph is d-regular. It satisfies* (A1), *potentially with a slightly different constant than G. However*, *the new edges create a bottleneck for the random walk*, *implying that the spectral gap of the new graph decreases to zero with the number of vertices n.*

The paper is organised as follows. In Section 2, we set up the notation. In Section 3, we prove an estimate of the form (1.6). The piecewise independent measure Q is constructed in Section 4. Sections 5, 6 and 7 contain the proofs of Theorems 1.1, 1.2 and 1.3, respectively.

2. Notation

In this section we introduce additional notation and recall some known results about random interlacements.

2.1. *Basic notations*

Throughout the text c or c' denote strictly positive constants only depending on d, and the parameters α_1 and α_2 in assumptions (A1) and (A2), with value changing from place to place. The numbered constants $c_0, c_1, \dots$ are fixed and refer to their first appearance in the text. Dependence of constants on additional parameters appears in the notation. For instance c_γ denotes a positive constant depending on γ and possibly on d, α_1, α_2.

We write $\mathbb{N} = \{0, 1, \dots\}$ for the set of natural numbers and for $a \in \mathbb{R}$ we write $\lfloor a \rfloor$ for the largest integer smaller or equal to a and define $\lceil a \rceil = \lfloor a \rfloor + 1$. In this paper we use $\ln x$ for the natural logarithm and use ld to denote the logarithm with base $d-1$,

$$\operatorname{ld} x = \log_{d-1} x = \ln x / \ln(d-1). \tag{2.1}$$

For a set A we denote by $|A|$ its cardinality.

Recall that we have introduced a sequence of finite connected graphs $G_k = (V_k, \mathcal{E}_k)$ in the Introduction. We will always omit the subscript k of the sequence of graphs G_k and their sizes n_k. In particular, we always assume that n is the number of vertices of G. For d as in (A0), we will also consider the infinite d-regular tree, denoted $\mathbb{T}_d = (\mathbb{V}_d, \mathbb{E}_d)$.

We now introduce some notation valid for an arbitrary graph $G = (V, \mathcal{E})$. We use $\operatorname{dist}(\cdot,\cdot)$ to denote the usual graph distance and write $x \sim y$, if x, y are neighbours in G. We write $B(x,r)$ for the ball centred at x with radius r, $B(x,r) = \{y \in V\colon \operatorname{dist}(x,y) \le r\}$. For $A \subset V$ we define its complement $A^c = V \setminus A$, its r-neighbourhood $B(A,r) = \bigcup_{x \in A} B(x,r)$, and its interior and exterior boundary

$$\partial_{\mathrm{i}} A = \big\{x \in A\colon \exists y \in A^c, x \sim y\big\}, \qquad \partial_{\mathrm{e}} A = \big\{x \in A^c\colon \exists y \in A, x \sim y\big\}. \tag{2.2}$$

We write $\operatorname{int}(A)$ for $A \setminus \partial_{\mathrm{i}} A$. We define the tree excess of a connected set $A \subset V$, denoted by $\mathtt{tx}(A)$, as the number of edges which can be removed from the subgraph of G induced by A while keeping it connected. Equivalently,

$$\mathtt{tx}(A) = |\mathcal{E}_A| - |A| + 1, \tag{2.3}$$

where $\mathcal{E}_A$ stands for the edges of the subgraph induced by A. By a *cycle* we mean a sequence of vertices $x_1, \dots, x_k$ such that $x_1 = x_k$ and $x_{i+1} \sim x_i$ for all $1 \le i < k$. Note that $\mathtt{tx}(A) = 0$ if and only of there is no cycle in A.

2.2. *Random walk on graphs*

We use P_x to denote the law of the canonical continuous-time simple random walk on G started at $x \in V$, that is of the Markov process with generator given by

$$\Delta f(x) = \sum_{y \in V} \big(f(y) - f(x)\big) p_{xy} \quad \text{for } f : V \to \mathbb{R}, x \in V, \tag{2.4}$$

where $p_{xy} = 1/d_x$ if $x \sim y$, and $p_{xy} = 0$ otherwise, and d_x denotes the degree of the vertex x. We write P_x^G for P_x whenever ambiguity would otherwise arise.

With exception of Lemma 3.1 and Proposition 3.2, we will always work with regular graphs G, in which case d_x is the same for every vertex $x \in V$. We use X_t to denote the canonical process and $(\mathcal{F}_t)_{t \ge 0}$ the canonical filtration. We write P_x^ℓ for the restriction of P_x to $D([0,\ell], V)$ and P_{xy}^ℓ for the law of random walk bridge, that is for P_x^ℓ conditioned on $X_\ell = y$. We write $E_x, E_x^\ell, E_{xy}^\ell$ for the corresponding expectations. The canonical shifts on $D([0,\infty), V)$ are

denoted by θ_t. The time of the nth jump is denoted by τ_n, i.e. $\tau_0 = 0$ and for $n \geq 1$, $\tau_n = \inf\{t \geq 0\colon X_t \neq X_0\} \circ \theta_{\tau_{n-1}} + \tau_{n-1}$. The process counting the number of jumps before time t is denoted by $N_t = \sup\{k\colon \tau_k \leq t\}$. Note that under P_x, $(N_t)_{t\geq 0}$ is a Poisson process on $\mathbb{R}_+$ with intensity 1, but this is not true under P^{ℓ}_{xy}. We write $\hat{X}_n$ for the discrete skeleton of the process X_t, that is $\hat{X}_n = X_{\tau_n}$. For $0 \leq s \leq t$, we use $X_{[s,t]}$ to denote the set of vertices visited by the random walk between times s and t, $X_{[s,t]} = \{X_r\colon s \leq r \leq t\}$.

Given $A \subset V$, we denote with H_A and $\tilde{H}_A$ the respective entrance and hitting time of A

$$H_A = \inf\{t \geq 0\colon X_t \in A\} \quad \text{and} \quad \tilde{H}_A = H_A \circ \theta_{\tau_1} + \tau_1. \tag{2.5}$$

We write $\hat{H}_A$ for the discretised entrance time, $\hat{H}_A = N_{H_A}$.

For the remaining notation, we assume that G is a finite connected graph. For such G we denote by π the stationary distribution for the simple random walk on G and use π_x for $\pi(x)$. P stands for the law of the simple random walk started at π and E for the corresponding expectation. Under assumption (A0) the stationary distribution is the uniform distribution. For all real valued functions f, g on V we define the Dirichlet form

$$\mathcal{D}(f,g) = \frac{1}{2}\sum_{x,y\in V}\big(f(x)-f(y)\big)\big(g(x)-g(y)\big)\pi_x p_{xy} = -\sum_{x\in G}\Delta f(x)g(x)\pi_x. \tag{2.6}$$

The spectral gap of G is given by

$$\lambda_G = \min\big\{\mathcal{D}(f,f)\colon \pi\big(f^2\big) = 1, \pi(f) = 0\big\}. \tag{2.7}$$

From [19], p. 328, it follows that under assumption (A0),

$$\sup_{x,y\in V}\big|P_x[X_t = y] - \pi_y\big| \leq \mathrm{e}^{-\lambda_G t} \quad \text{for all } t \geq 0. \tag{2.8}$$

A function $h\colon V \to \mathbb{R}$ is called harmonic on A if $\Delta h(x) = 0$ for all $x \in A$. For two non-empty disjoint subsets A, C of V we define the equilibrium potential $g^{\star}_{A,C}$ as the unique function harmonic on $(A \cup C)^c$, satisfying $g^{\star}|_A = 1$, $g^{\star}|_C = 0$. It is well known that

$$g^{\star}_{A,C}(x) = P_x[H_A \leq H_C], \tag{2.9}$$

$$\mathcal{D}\big(g^{\star}_{A,C}, g^{\star}_{A,C}\big) = \sum_{z\in A} P_z[\tilde{H}_A > H_C]\pi_z. \tag{2.10}$$

We define the isoperimetric constant of G as $\iota_G = \min\{|\partial_{\mathrm{e}} A|/|A|\colon A \subset V, |A| \leq |V|/2\}$. If assumption (A0) holds, then Cheeger's inequality ([19], Lemma 3.3.7) yields $c\iota_G^2 \leq \lambda_G \leq c'\iota_G$. The assumption (A2) then implies the existence of $\alpha_2' > 0$ such that

$$|\partial_{\mathrm{e}} A| \geq \alpha_2'|A| \quad \text{for all } k \geq 1 \text{ and } A \subset V \text{ with } |A| \leq |V|/2. \tag{2.11}$$

2.3. *Random interlacement*

Let us give a brief introduction to random interlacements. Although we will not directly use any results on random interlacements in this paper, random interlacements give a natural interpretation to the key result in Section 6. Consider an infinite locally finite graph $\mathbb{G} = (\mathbb{V}, \mathbb{E})$ for which the simple random walk (with law denoted by $P^{\mathbb{G}}_x$) is transient. According to [25,26], the interlacement set on $\mathbb{G}$ is given by the trace left by a Poisson point process of doubly infinite trajectories modulo time-shift in $\mathbb{G}$ which visit every point only finitely many times. The complement of the interlacement set is called vacant set. Although the precise construction of the random interlacements on a graph is delicate, we give here a characterization of the law $\mathbb{Q}_u$ that the vacant set induces on $\{0,1\}^{\mathbb{V}}$. For this, consider a finite set $K \subset \mathbb{V}$ and define the capacity of K as

$$\mathrm{cap}(K) = \sum_{x\in K} P^{\mathbb{G}}_x[\tilde{H}_K = \infty] \tag{2.12}$$

with $\tilde{H}_K$ as in (2.5). The law $\mathbb{Q}_u$ of the indicator function of the vacant set at level u is the only measure on $\{0,1\}^{\mathbb{V}}$ such that

$$\mathbb{Q}_u[W_y = 1 \text{ for all } y \in K] = \exp\{-u \cdot \operatorname{cap}(K)\}, \tag{2.13}$$

where $\{W_y\}_{y\in\mathbb{V}}$ are the canonical projections from $\{0,1\}^{\mathbb{V}}$ to $\{0,1\}$, see (1.1) in [26].

3. Conditional probability estimate

In this section, we derive in Proposition 3.6 an estimate on the probability that in a finite time interval $[0,T]$, a random walk does not visit a vertex y in the boundary of a set A, given that it does not visit the set A. This estimate will be crucial in the analysis of the breadth-first search algorithm exploring the components of the vacant set, used in the proof of Theorem 1.1.

We recall first a variational formula for the expected entrance time.

Lemma 3.1 ([2], Chapter 3, Proposition 41). *For a non-empty subset $A \subseteq V$,*

$$(EH_A)^{-1} = \inf\{\mathcal{D}(f,f)\colon\ f : V \to \mathbb{R},\ f = 1 \text{ on } A, \pi(f) = 0\}. \tag{3.1}$$

The minimizing function $f^\star$ in (3.1) is given by

$$f^\star(x) = 1 - \frac{E_x H_A}{EH_A}. \tag{3.2}$$

Using this variational formula, we obtain the following estimate.

Proposition 3.2. *Let A and C be disjoint non-empty subsets of V and let $g^\star = g^\star_{A,C}$ be the equilibrium potential (2.9) and $f^\star$ the minimising function (3.2). Then*

$$\mathcal{D}(g^\star, g^\star)\Big(1 - 2\sup_{x\in C}|f^\star(x)|\Big) \le \frac{1}{E[H_A]} \le \mathcal{D}(g^\star, g^\star)\pi(C)^{-2}. \tag{3.3}$$

Proof. We prove the right-hand inequality in (3.3) first. To this end, we modify the function $g^\star$ such that it becomes admissible for the variational problem (3.1) of Lemma 3.1. We define the function $\tilde{g}$ on V by $\tilde{g} = (g^\star - \pi(g^\star))/(1-\pi(g^\star))$. Then $\tilde{g}$ equals 1 on A and $\pi(\tilde{g}) = 0$, so we obtain from (3.1) that

$$E[H_A]^{-1} \le \mathcal{D}(\tilde{g},\tilde{g}) = \mathcal{D}(g^\star, g^\star)\big(1 - \pi(g^\star)\big)^{-2}. \tag{3.4}$$

Since $g^\star$ is non-negative, bounded by 1 and non-zero only on C^c, we have $\pi(g^\star) \le \pi(C^c)$ and the right-hand inequality of (3.3) follows.

To prove the left-hand inequality in (3.3), observe that the minimiser $f^\star$ of the variational problem (3.1) satisfies $f^\star = 1$ on A. Therefore

$$E[H_A]^{-1} = \mathcal{D}(f^\star, f^\star) \ge \inf\{\mathcal{D}(g,g)\colon\ g : V \to \mathbb{R}, g = 1 \text{ on } A, g = f^\star \text{ on } C\}. \tag{3.5}$$

Since G is finite, the infimum is attained by a function $\hat{g}$ which satisfies the given boundary conditions on A and C and which is harmonic in $(A\cup C)^c$. In particular, the process $(\hat{g}(X_{t\wedge H_{A\cup C}}))_{t\ge 0}$ is a P_x-martingale for any $x \in V$. From the optional stopping theorem, it follows that $\hat{g}(x) = g^\star(x) + \psi(x)$, $x \in V$, where the function ψ is defined by

$$\psi(x) = E_x\big[f^\star(X_{H_C})\mathbf{1}_{\{H_C < H_A\}}\big] \quad \text{for } x \in V. \tag{3.6}$$

Therefore,

$$E[H_A]^{-1} \ge \mathcal{D}(g^\star + \psi, g^\star + \psi) \ge \mathcal{D}(g^\star, g^\star) + 2\mathcal{D}(g^\star, \psi). \tag{3.7}$$

Since ψ equals 0 on A, $\Delta g^\star$ equals 0 on $(A \cup C)^c$ and on $\operatorname{int} C$, and $\Delta g^\star(x) \geq 0$ for all $x \in \partial_{\mathrm{i}} C$ (indeed, $g^\star$ is non-negative on V and equal to 0 on $\partial_{\mathrm{i}} C$), we have

$$\mathcal{D}\big(g^\star, \psi\big) = -\sum_{x \in \partial_{\mathrm{i}} C} \Delta g^\star(x)\psi(x)\pi_x \geq -|\psi|_\infty \sum_{x \in \partial_{\mathrm{i}} C} \Delta g^\star(x)\pi_x. \tag{3.8}$$

Using again $g^\star$ equals 0 on $\partial_{\mathrm{i}} C$ and observing that $\Delta f = -\Delta(1-f)$ for any real-valued function f on V, as can be directly seen from the definition of Δ in (2.4), we obtain

$$\mathcal{D}\big(g^\star, \psi\big) \geq |\psi|_\infty \sum_{x \in \partial_{\mathrm{i}} C} \Delta\big(1-g^\star\big)(x)\big(1-g^\star(x)\big)\pi_x. \tag{3.9}$$

Since $1-g^\star$ vanishes on A, while $\Delta(1-g^\star) = \Delta g^\star$ vanishes on $(A \cup C)^c$ and on $\operatorname{int} C$, the right-hand side equals $-|\psi|_\infty \mathcal{D}(1-g^\star, 1-g^\star) = -|\psi|_\infty \mathcal{D}(g^\star, g^\star)$. Putting together (3.7) and (3.9) and using (3.6), we therefore obtain

$$E[H_A]^{-1} \geq \mathcal{D}\big(g^\star, g^\star\big)\big(1-2|\psi|_\infty\big) \geq \mathcal{D}\big(g^\star, g^\star\big)\Big(1-2\sup_{x \in C}\big|f^\star(x)\big|\Big).$$

This yields the left-hand estimate in (3.3) and completes the proof of Proposition 3.2. □

In order to apply the left-hand estimate of (3.3), a bound on $\sup_{x \in C} |f^\star(x)|$ is required. We will derive such a bound in Proposition 3.5 below. In its proof we will need the following technical lemma.

Lemma 3.3. *Assume* (A0) *and consider* $r, s \in \mathbb{N}$ *and* $x \in V$, *such that* $\mathtt{tx}(B(x, r+s)) \leq 1$. *Then for any* $y \in \partial_{\mathrm{i}} B(x, r+s)$,

$$P_y[H_{B(x,r)} < H_{B(x,r+s)^c}] \leq c(d-1)^{-s}. \tag{3.10}$$

Proof. We write $B = B(x, r)$ and $B' = B(x, r+s)$ and for every vertex $z \in B'$ we define $r_z = \operatorname{dist}(x, z)$. If $\mathtt{tx}(B') = 0$, then B' is a tree and r_{X_t} behaves like a random walk on $\mathbb{N}$ with drift, which steps left with probability $p = 1/d$ and right otherwise.

It is a known fact that the probability that a random walk on $\mathbb{Z}$ jumping with probability p to the right and $1-p$ to the left started at $x > 0$ hits $R \geq x$ before hitting zero equals (see, e.g., [10], Chapter 4, Example 7.1)

$$\frac{q^x - 1}{q^R - 1}, \quad \text{where } q = (1-p)/p. \tag{3.11}$$

The inequality (3.10) then follows directly from (3.11).

We thus assume that $\mathtt{tx}(B') = 1$. Let us call a vertex z in $B' \setminus \{x\}$ exceptional, if z does not have $d-1$ neighbours z' with $r_{z'} > r_z$. We claim that

There are at most two exceptional vertices. All of them are at the same distance (say ρ) of x and have at most two neighbours z' with $r_{z'} \leq r_z$. (3.12)

To see this, consider an exceptional vertex $z \in B'$. By definition, there is a pair z_1, z_2 of neighbours of z with $r_{z_1}, r_{z_2} \leq r_z$. By considering geodesic paths from z_1 and z_2 to x, one can extract a cycle in $B(x, r_z)$ containing z and exactly two of its neighbours z_1, z_2. By construction, this cycle has at most two vertices which maximize the distance to x. One of them is z. Second might be z_1 or z_2, in which case this vertex has the same distance to x as z and is also exceptional. To show that there cannot be another exceptional vertex other than z (and potentially one of z_1, z_2), we suppose that there is one, we call it z'. By the same reasoning we can extract a cycle in B' containing z' with z' maximizing the distance to x. This cycle thus must be different from the one containing z. This is impossible since $\mathtt{tx}(B') = 1$. Similarly, if z has three or more neighbours z_i with $r_{z_i} \leq r_z$, then every pair of them can be used to extract a cycle, all of them being different. This is again in contradiction with $\mathtt{tx}(B') = 1$. With this we conclude (3.12).

Let $Y_t = \mathrm{dist}(B, X_{t\wedge H_{B\cup(B')^c}})$. We compare Y with a continuous-time birth–death process U_t on $\{0,\dots,s+1\}$ given by the following transition rates

$$p_{i,i+1} = 1 - p_{i,i-1} = \begin{cases} (d-2)/d, & \text{if } i = \rho - r, \\ (d-1)/d, & \text{if } i \in \{1,\dots,s\} \setminus \{\rho - r\}, \end{cases} \tag{3.13}$$

and such that the states 0 and $s+1$ are absorbing. More precisely, using (3.12), we can couple Y (under law P_y) with U (started from s) in such way that $U_t \le Y_k$ for every $t \ge 0$. This implies that

$$P_y[H_B < H_{(B')^c}] \text{ is smaller or equal to the probability that } U \text{ hits 0 before } s+1, \text{ given that } U_0 = s. \tag{3.14}$$

The last probability will now be estimated using a standard birth–death process computation. Let $f(i)$ be the probability that U started at i hits 0 and set $h(i) = f(i-1) - f(i)$, $i \in \{1,\dots,s+1\}$. Clearly $f(0) = 1$, $f(s+1) = 0$ and the strong Markov property on the time of the first jump implies that $h(i)p_{i,i-1} = h(i+1)p_{i,i+1}$, $i \in \{1,\dots,s\}$. Fixing $h(s+1) = \gamma$ we use the above facts to get

$$1 = f(0) \ge h(1) = \gamma \cdot \frac{p_{1,2}\cdots p_{s,s+1}}{p_{1,0}\cdots p_{s,s-1}} = \frac{\gamma}{2}(d-1)^{s-1}(d-2)2. \tag{3.15}$$

Moreover, conditioned on $U_0 = s$, the probability that U hits zero before $s+1$ is $f(s) = \gamma$, i.e. Prob[U hits 0 before $s+1 | U_0 = s] = f(s) = \gamma$. Putting this together with (3.14) and (3.15) the proof of Lemma 3.3 is finished. □

We now apply the last lemma to estimate the probability that the random walk, started outside of the larger one of two concentric balls visits the small ball before time $T > 0$.

Lemma 3.4. *Assume that G satisfies* (A0) *and consider $T > 0$, $r, s \in \mathbb{N}$ and $x \in V$ such that* $\mathtt{tx}(B(x, r+s)) \le 1$. *Then, for some $c, c' > 0$,*

$$P_y[H_{B(x,r)} < T] \le cT(d-1)^{-s} + \mathrm{e}^{-c'T} \quad \textit{for all } y \in B(x, r+s)^c. \tag{3.16}$$

Proof. As in the previous proof we write $B = B(x,r)$, $B' = B(x, r+s)$. From an exponential upper bound on the probability that a Poisson random variable with expectation T is larger than $2T$, we have

$$P_y[H_B < T] \le P_y[\hat{H}_B \le 2T] + \mathrm{e}^{-c'T}, \tag{3.17}$$

where $\hat{H}_B$ is the entrance time for the discrete-time walk defined below (2.5).

On the way from y to x (as in the lemma), the simple random walk must visit some vertex $z \in \partial_{\mathrm{i}} B'$. After reaching such vertex, it either hits B without exiting B' or it exits B'. The probability of the first event is bounded from above by $c(d-1)^{-s}$, see Lemma 3.3. When the second event occurs, the simple random walk must again pass through $\partial_{\mathrm{i}} B'$ in order to visit x. At this point we can repeat the previous reasoning. However, before time $2T$ we can repeat this procedure at most $2T$ times, since we are considering a discrete-time walk. A union bound then implies

$$P_y[\hat{H}_B \le 2T] \le 2Tc(d-1)^{-s} \tag{3.18}$$

and Lemma 3.4 follows by renaming constants. □

Finally, we prove the proposition that will allow us to use the left-hand side of the estimate (3.3) on $E[H_A]^{-1}$, derived in the beginning of this section.

Proposition 3.5. *Let $G = (V, \mathcal{E})$ be a graph on n vertices satisfying* (A0), (A2) *and let $A \subseteq V$, $s \in (0, \log_{d-1} n] \cap \mathbb{N}$ such that $|A| \le n/2$ and* $\mathtt{tx}(B(x,s)) \le 1$ *for every $x \in A$. Then*

$$\sup_{y \in V : \mathrm{dist}(y,A) > s} \left| \frac{E_y[H_A]}{E[H_A]} - 1 \right| \le c|A|(d-1)^{-s} \log^4 n. \tag{3.19}$$

Proof. In essence, the proof is an application of the estimate (2.8), which shows that the distribution of the random walk on G at time $T = \lambda_G^{-1} \log^2 n$ is close to uniform. From Lemma 3.4, we know that it is unlikely that the random walk started at y reaches a point x in A before time T and this will yield (3.19).

We shall require the following rough bounds:

$$\frac{n}{4|A|} \le E[H_A] \le \sup_{z\in G} E_z[H_A] \le cn \log n \quad \text{for some constant } c > 0. \tag{3.20}$$

The first inequality in (3.20) follows from the right-hand estimate of (3.3) with C chosen as A^c, (2.10), and our assumption that $|A| \le n/2$. To prove the last inequality in (3.20), observe that for $t = 2\log n/\alpha_2$, assumption (A2) and (2.8) imply

$$\inf_{z\in V} P_z[H_A \le 2\log n/\alpha_2] \ge \inf_{z\in V} P_z[X_{2\log n/\alpha_2} \in A] \ge (2n)^{-1}. \tag{3.21}$$

By the simple Markov property applied at integer multiples of t, it follows that H_A is stochastically dominated by t times a geometrically distributed random variable with success probability $1/2n$ and (3.20) readily follows.

Let y be chosen as in the statement and let us first consider the expectation of H_A starting from X_T. From (2.8) and our crude estimate (3.20), we obtain, for any $z \in V$,

$$\begin{aligned} \bigl|E_z\bigl[E_{X_T}[H_A]\bigr] - E[H_A]\bigr| &\le \sum_{z'\in V} \bigl|P_z\bigl[X_T = z'\bigr] - \pi_{z'}\bigr| E_{z'}[H_A] \\ &\le \sum_{z'\in V} \mathrm{e}^{-\log^2 n} n \log n \le n^3 \mathrm{e}^{-\log^2 n}. \end{aligned} \tag{3.22}$$

We now apply this inequality to find an upper bound on $E_y[H_A]$. Since $H_A \le T + H_A \circ \theta_T$, the simple Markov property applied at time T and (3.22) imply that for any $z \in V$,

$$E_z[H_A] \le T + E_z\bigl[E_{X_T}[H_A]\bigr] \le T + n^3 \mathrm{e}^{-\log^2 n} + E[H_A]. \tag{3.23}$$

With the first inequality in (3.20), we deduce that

$$\frac{E_z[H_A]}{E[H_A]} - 1 \le \bigl(T + n^3 \mathrm{e}^{-\log^2 n}\bigr) \frac{4|A|}{n} \le \frac{c|A|\log^2 n}{n}, \tag{3.24}$$

which is ample for one side of (3.19). To prove the other half of (3.19), choose y as in the statement and apply the simple Markov property at time T to infer that

$$\begin{aligned} E_y[H_A] &\ge E_y\bigl[\mathbf{1}_{\{H_A > T\}} E_{X_T}[H_A]\bigr] = E_y\bigl[E_{X_T}[H_A]\bigr] - E_y\bigl[\mathbf{1}_{\{H_A \le T\}} E_{X_T}[H_A]\bigr] \\ &\overset{(3.22)}{\ge} E[H_A] - n^3 \mathrm{e}^{-\log^2 n} - P_y[H_A \le T] \sup_{z\in V} E_z[H_A] \\ &\overset{(3.23)}{\ge} E[H_A] - 2n^3 \mathrm{e}^{-\log^2 n} - P_y[H_A \le T]\bigl(T + E[H_A]\bigr). \end{aligned} \tag{3.25}$$

Applying (3.16) to the probability on the right-hand side and rearranging, we find that

$$\frac{E_y[H_A]}{E[H_A]} - 1 \ge -c|A|(d-1)^{-s} \log^4 n, \tag{3.26}$$

which together with (3.24) completes the proof of Proposition 3.5. □

We now analyse the distribution of the hitting time of a point y conditioned on the event that a certain set A is vacant. This estimate will be helpful for the analysis of the breadth-first search algorithm used in Theorem 1.1.

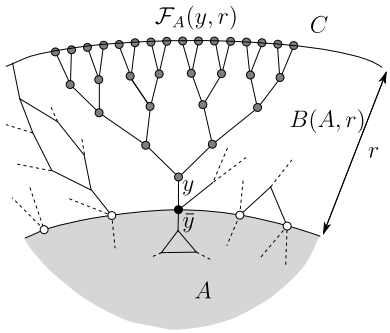

Fig. 1. Proper $\mathcal{F}_A(y,r)$ (gray points) on a 3-regular graph with $r=5$.

For any non-empty connected set $A \subset V$, $r \geq 1$ and $y \in \partial_{\mathrm{e}} A$ we define

$$\mathcal{F}_A(y,r) = \big\{z \in B(A,r) \setminus A\colon\ z \text{ is connected to } y \text{ in } B(A,r) \setminus A\big\}. \tag{3.27}$$

Observe that $y \in \mathcal{F}_A(y,r)$. In the breadth-first search algorithm to be introduced in Section 5, the set $\mathcal{F}_A(y,r)$ can be viewed as the 'future of y seen from A.' We say that $\mathcal{F}_A(y,r)$ is *proper* when (see Fig. 1)

$$\begin{aligned}
&\text{(i)} \quad \mathtt{tx}(\mathcal{F}_A(y,r)) = 0,\\
&\text{(ii)} \quad y \text{ has a unique neighbour } \bar{y} \text{ in } A,\\
&\text{(iii)} \quad \text{for any vertex } y' \in A \setminus \bar{y}, \text{ every path from } y \text{ to } y' \text{ leaves } B(A,r) \setminus A \text{ before reaching } y'.
\end{aligned} \tag{3.28}$$

Proposition 3.6. *Let $s \in [2, (\alpha_1 \wedge \frac{1}{2}) \ln n)$, $A \subset V$, $A \neq \varnothing$ with $|B(A,s)| \leq \sqrt{n}$, and $y \in \partial_{\mathrm{e}} A$, such that $\mathcal{F}_A(y,s)$ is proper. Then, for any $T > 1$,*

$$\left| \ln P[H_{A\cup\{y\}} > T \,|\, H_A > T] + \frac{T}{n} \frac{(d-2)^2}{d(d-1)} \right| \leq \frac{c|A|}{n} \left(\frac{T|A| \ln^4 n}{(d-1)^s} + 1 \right). \tag{3.29}$$

Proof. We set

$$F_A^y(T) = P[H_{A\cup\{y\}} > T \,|\, H_A > T] = \frac{P[H_{A\cup\{y\}} > T]}{P[H_A > T]}, \tag{3.30}$$

and use results of [1] to estimate both numerator and denominator. Namely, by [1], (1) and Theorem 3, for any $A \subset V$, $t > 0$,

$$\left(1 - \frac{1}{\lambda_G E_{\alpha_A} H_A}\right) \exp\left(-\frac{t}{E_{\alpha_A} H_A}\right) \leq P[H_A > t] \leq \big(1 - \pi(A)\big) \exp\left(-\frac{t}{E_{\alpha_A} H_A}\right). \tag{3.31}$$

Here α_A is the quasi-stationary distribution for the random walk killed on hitting A. We will only need its following properties, see [1], Lemma 2 and Corollary 4,

$$\frac{1-\pi(A)}{\sum_{x\in A, y\in A^c} \pi(x) p_{xy}} \leq \frac{E H_A}{1-\pi(A)} \leq E_{\alpha_A} H_A \leq E H_A + \lambda_G^{-1}. \tag{3.32}$$

Observe that the left-hand side is bounded from below by $n/(2|A|) \geq c\sqrt{n}$ for A as in the statement.

Writing $\tilde{A} = A \cup \{y\}$, $\alpha = \alpha_A$ and $\tilde{\alpha} = \alpha_{\tilde{A}}$, and applying (3.31) for A as well as for $\tilde{A}$ to bound the conditional expectation (3.30), we obtain after rearranging and taking logarithm

$$\ln \frac{1 - 1/(\lambda_G E_{\tilde{\alpha}} H_{\tilde{A}})}{1-\pi(A)} \leq \ln F_A^y(T) - \frac{T}{E_\alpha H_A} + \frac{T}{E_{\tilde{\alpha}} H_{\tilde{A}}} \leq \ln \frac{1-\pi(\tilde{A})}{1 - 1/(\lambda_G E_\alpha H_A)}. \tag{3.33}$$

Using (3.32) and the observation following it, we see that $E_\alpha H_A \geq c\sqrt{n}$. Therefore, by expanding the function $\ln(1/(1-x))$ around zero and using (3.32) again, we obtain that the right most term in (3.33) is bounded from above by $c|A|/n$, which can be included in the error term of (3.29). A similar reasoning implies that the left-most term in (3.33) is bounded from below by $-c|A|/n$, which can again be accommodated into the error of (3.29).

The inequalities in (3.32) further imply that $0 \leq E_\alpha H_A - EH_A \leq \lambda_G^{-1}$ and therefore

$$\left|\frac{T}{E_\alpha H_A} - \frac{T}{EH_A}\right| \leq \frac{cT}{(EH_A)^2} \leq cT\frac{|A|^2}{n^2}, \tag{3.34}$$

where in the last inequality we used (3.32) again. The right-hand term in the last display is again smaller than the error in (3.29), since $(d-1)^{-s} \geq n^{-1/2}$ by assumption $s \leq \frac{1}{2}\operatorname{ld} n$.

Finally, we use Proposition 3.2 to approximate $1/EH_A$ and $1/EH_{\tilde{A}}$. To this end we introduce $C = B(A,s)^c$ and we set $g = g^\star_{A,C}$, see (2.9). Then by Proposition 3.2 we have that

$$\left|\frac{T}{EH_A} - T\mathcal{D}(g,g)\right| \leq T\mathcal{D}(g,g)\left(\pi(C)^{-2} - 1 + 2\sup_{z\in C}\left|\frac{E_zH_A}{EH_A} - 1\right|\right). \tag{3.35}$$

By (2.10), $\mathcal{D}(g,g) \leq \pi(A)$. From the assumption $B(A,s) \leq \sqrt{n}$ it follows that $\pi(C)^{-2} - 1 \leq cn^{-1/2} \leq c(d-1)^{-s}$. Finally, the Proposition 3.5 implies that the supremum in (3.35) is bounded by $c|A|(d-1)^{-s}\ln^4 n$. Hence, the right-hand side of (3.35) is smaller than the error term in (3.29). An analogous computation proves that $T/EH_{\tilde{A}}$ is approximated by $T\mathcal{D}(\tilde{g},\tilde{g})$, where $\tilde{g} = g^\star_{\tilde{A},C}$. A little bit of care is only needed when applying Proposition 3.5, since $\operatorname{dist}(\tilde{A}, C) = s-1$.

We have thus proved that $\ln F_A^y(T)$ is well approximated by $T(\mathcal{D}(g,g) - \mathcal{D}(\tilde{g},\tilde{g}))$ up to the error on the right-hand side of (3.29). We now estimate this expression. Let $\bar{y}$ be the unique neighbour of y in A. By (2.10),

$$T\big(\mathcal{D}(g,g) - \mathcal{D}(\tilde{g},\tilde{g})\big) = \frac{T}{n}\left[\sum_{z\in A} P_z[\tilde{H}_A > H_C] - \sum_{z\in\tilde{A}} P_z[\tilde{H}_{\tilde{A}} > H_C]\right]. \tag{3.36}$$

Now we use our assumption that $\mathcal{F}_A(y,s)$ is proper. Due to (3.28)(iii), for all $z \in A\setminus\{\bar{y}\}$, there is no path from z to y using only vertices in $B(A,r)\setminus A$. Therefore, for such z $P_z[\tilde{H}_{\tilde{A}} > H_C] = P_z[\tilde{H}_A > H_C]$, and (3.36) equals

$$\frac{T}{n}\big[P_{\bar{y}}[\tilde{H}_A > H_C] - P_{\bar{y}}[\tilde{H}_{\tilde{A}} > H_C] - P_y[\tilde{H}_{\tilde{A}} > H_C]\big]. \tag{3.37}$$

Conditioning the first two terms on X_{τ_1}, since $\mathcal{F}_A(y,s)$ is proper, we get

$$\frac{T}{n}\left\{\frac{1}{d}P_y[H_A > H_C] - P_y[\tilde{H}_{\tilde{A}} > H_C]\right\}. \tag{3.38}$$

Since by assumption $\mathtt{tx}(\mathcal{F}_A(y,s)) = 0$, these probabilities can be computed using the formula (3.11) for the random walk with drift. Setting $q = 1/(d-1)$ we have

$$P_y[\tilde{H}_{\tilde{A}} > H_C] = \frac{d-1}{d}\frac{1-q}{1-q^{s-1}} \quad \text{and} \quad P_y[H_A > H_C] = \frac{1-q}{1-q^s}. \tag{3.39}$$

Inserting this into (3.38) we obtain that

$$\left|T\big(\mathcal{D}(g,g) - \mathcal{D}(\tilde{g},\tilde{g})\big) + \frac{T(d-2)^2}{nd(d-1)}\right| \leq \frac{cT}{n}\frac{1}{(d-1)^s}. \tag{3.40}$$

This completes the proof, since the error is smaller than the right-hand side of (3.29). □

We now use the same techniques to control the hitting time distribution of a point with tree-like neighbourhood.

Lemma 3.7. *Let $y \in V$ be such that $\mathtt{tx}(B(y,s)) = 0$ for some $s \in [1, \alpha_1 \operatorname{ld} n)$. Then, for any $T > 1$,*

$$\left| \ln P[H_y > T] + \frac{T(d-2)}{n(d-1)} \right| \le \frac{c}{n} \left(\frac{T \ln^4 n}{(d-1)^s} + 1 \right). \tag{3.41}$$

Proof. The proof follows the same lines as the previous one with $\tilde{A} = \{y\}$ and without the conditioning, which is equivalent to controlling the numerator of (3.30) only. The same reasoning as before implies that, for $g = g^\star_{y,B(y,s)^c}$, $|\ln P[H_y \ge T] - T\mathcal{D}(g,g)|$ is smaller than the right-hand side of (3.41). Using (2.10), (3.11), with $q = 1/(d-1)$ again, we get

$$\mathcal{D}(g,g) = n^{-1} P_y[\tilde{H}_y > H_{B(y,s)^c}] = \frac{1}{n} \frac{1-q}{1-q^s} = \frac{d-2}{n(d-1)} + \mathrm{O}\big((d-1)^{-s}\big), \tag{3.42}$$

which finishes the proof. □

4. Piecewise independent measure

We now make another preparative step in order to prove our main results. In later sections, it will be convenient to split the random walk trajectory $X_{[0,un]}$ into smaller pieces and to treat pieces that are sufficiently distant in time as being independent of one another. Although this kind of independence does not hold under the random walk measure P^{un}, we will in this section construct a new measure on the space of trajectories with the desired independence properties. In Lemma 4.1, we then estimate the error we make when replacing P^{un} by this new measure.

For the construction, we choose real parameters

$$L = n^\gamma, \qquad \ell = (\ln n)^2, \tag{4.1}$$

where $\gamma \in (0,1)$ will be fixed later. We consider an abstract probability space (Ω, P) (expectation denoted by E) on which we define a sequence of i.i.d. random variables Y^i, $i \ge 0$, with values in $D([0,L],V)$ and the marginal distribution P^L (defined in Section 2.2). We set a_i, b_i to be the start- and the end-point of Y^i, $a_i = Y^i_0$, $b_i = Y^i_L$. On the same space Ω we further define a sequence of random variables Z^i, $i \ge 0$, with values in $D([0,\ell],V)$. Given a_i, b_i, $i \ge 0$, the random variables Z^i are independent, conditionally independent of the sequence (Y^i), and the random variable Z^i has the random-walk bridge distribution $P^\ell_{b_i,a_{i+1}}$. We call the Y^i's *segments* and Z^i's *bridges*.

We now concatenate the Y^i's and Z^i's to obtain an element of $D([0,\infty],V)$. More precisely, we define the concatenation mapping $\mathcal{X}$ from Ω to $D([0,\infty),V)$ as follows: For $t \ge 0$, let $i_t \in \mathbb{N}$ and $s \in [0, L+\ell)$ be given by $t = i_t(L+\ell) + s_t$. Then, for $t \ge 0$,

$$\mathcal{X}_t\big(Y^0, Z^0, Y^1, Z^1, \dots\big) = \begin{cases} Y^{i_t}_{s_t}, & \text{if } 0 \le s_t \le L, \\ Z^{i_t}_{s_t - L}, & \text{if } L < s_t < L+\ell. \end{cases} \tag{4.2}$$

The mapping $\mathcal{X}$ induces a new probability measure $Q = \mathsf{P} \circ \mathcal{X}^{-1}$ on $D([0,\infty),V)$. We use Q^s, $s \ge 0$, to denote the restriction of Q to $D([0,s],V)$. The measure $Q^\cdot$ will be used to approximate $P^\cdot$ later. We control this approximation now.

Lemma 4.1. *For every fixed $u > 0$, the measures P^{un} and Q^{un} are absolutely continuous and there exist constants $c, c' > 0$ depending only on α_2 such that*

$$\left| \frac{\mathrm{d}P^{un}}{\mathrm{d}Q^{un}} - 1 \right| \le c' u \mathrm{e}^{-c \ln^2 n}. \tag{4.3}$$

Proof. Let $u' \ge u$ be the smallest number such that $u'n$ is an integer multiple of $(L+\ell)$, and set $m = u'n/(L+\ell) \in \mathbb{N}$. Let further A be an arbitrary $\mathcal{F}_{un}$-measurable subset of $D([0,u'n],V)$. Since P^{un} and Q^{un} are the restrictions of $P^{u'n}$

and $Q^{u'n}$ to $D([0,un],V)$, it is sufficient to prove the lemma with u replaced by u'. To this end, we set $t_{2k}=k(L+\ell)$, $t_{2k+1}=k(L+\ell)+L$ for $k\in\{0,\dots,m\}$, and write

$$P^{u'n}[A]=\sum_{x_0,\dots,x_{2m}\in V}P^{u'n}[A|X_{t_i}=x_i,0\le i\le 2m]P^{u'n}[X_{t_i}=x_i,0\le i\le 2m]. \tag{4.4}$$

By the Markov property

$$P^{u'n}[X_{t_i}=x_i,0\le i\le 2m]=\pi(x_0)\prod_{k=0}^{m-1}P^L_{x_{2k}}[X_L=x_{2k+1}]P^\ell_{x_{2k+1}}[X_\ell=x_{2k+2}]. \tag{4.5}$$

The construction of the measure Q implies that

$$Q^{u'n}[A|X_{t_i}=x_i,0\le i\le 2m]=P^{u'n}[A|X_{t_i}=x_i,0\le i\le 2m], \tag{4.6}$$

$$Q^{u'n}[X_{t_i}=x_i,0\le i\le 2m]=\pi(x_0)\prod_{k=0}^{m-1}P^L_{x_{2k}}[X_L=x_{2k+1}]\pi(x_{2k+2}). \tag{4.7}$$

Comparing (4.5) and (4.7), it remains to control the ratio $P^\ell_x[X_\ell=y]/\pi(y)$. However, by (A2) and (2.8), this ratio is bounded by $1+n\mathrm{e}^{-\alpha_2\ell}$. Hence, (4.4) is bounded from above by

$$\begin{aligned}&\bigl(1+n\mathrm{e}^{-\alpha_2\ell}\bigr)^m\sum_{x_0,\dots,x_{2m}\in V}Q^{u'n}[A|X_{t_i}=x_i,0\le i\le 2m]Q^{u'n}[X_{t_i}=x_i,0\le i\le 2m]\\&\quad\le Q^{u'n}[A]\bigl(1+u\mathrm{e}^{-c\ln^2 n}\bigr),\end{aligned} \tag{4.8}$$

where, in the last inequality, we changed the constants to accommodate the terms polynomial in n. A lower bound can be obtained analogously. We have thus shown

$$Q^{u'n}[A]\bigl(1-u\mathrm{e}^{-c\ln^2 n}\bigr)\le P^{u'n}[A]\le Q^{u'n}[A]\bigl(1+u\mathrm{e}^{-c\ln^2 n}\bigr). \tag{4.9}$$

It immediately follows that $P^{u'n}$ and $Q^{u'n}$ are absolutely continuous. Moreover, the fact that (4.9) holds for any event A in $\mathcal{F}_{u'n}$ yields directly the estimate (4.3). □

We end this section with a simple lemma which controls the number of jumps performed by segments and bridges, which will be useful several times later (see Section 2.2 for the definition of the jump process $(N_t)_{t\ge 0}$).

Lemma 4.2.

$$P\bigl[2^{-1}n^\gamma<N_L<2n^\gamma\bigr]\ge 1-\mathrm{e}^{-c_\gamma n^\gamma}. \tag{4.10}$$

For any $x,y\in V$,

$$P^\ell_{xy}\bigl[N_\ell>\ln^3 n\bigr]\le c\exp\bigl\{-c'\ln^3 n\bigr\}. \tag{4.11}$$

Proof. Under the measure P, the random variable N_L has Poisson distribution with parameter $L=n^\gamma$. Hence, (4.10) follows by a standard large deviation argument.

In order to prove (4.11), note first that N_ℓ is not necessarily a Poisson random variable under P^ℓ_{xy}, due to the conditioning on the position of the endpoint. However, using (2.8) for the last inequality, we have

$$\begin{aligned}\sup_{x,y\in V}E^\ell_{x,y}\bigl[\mathrm{e}^{N_\ell}\bigr]&=\sup_{x,y\in V}E^\ell_x\bigl[\mathrm{e}^{N_\ell}|X_\ell=y\bigr]\le\frac{\sup_{x\in V}E^\ell_x[\mathrm{e}^{N_\ell}]}{\inf_{x,y\in V}P_x[X_\ell=y]}\\&\le\frac{\exp\{(\mathrm{e}-1)\ell\}}{\inf_{x,y\in V}P_x[X_\ell=y]}\le 2n^{-1}\exp\bigl\{c\ln^2 n\bigr\}\end{aligned} \tag{4.12}$$

for n larger than some c'. The exponential Chebyshev inequality then implies claim (ii) for such n. Adjusting the constants to make the claim valid for all n finishes the proof. □

5. Sub-critical regime

In this section we prove Theorem 1.1, which states that if $u > u_\star$, then the maximal connected component $\mathcal{C}_{\max}$ of $\mathcal{V}_n^u$ is typically of size $\mathrm{O}(\ln n)$. We will do it by analysing a breadth-first-search (BFS) algorithm which explores the component $\mathcal{C}_x$ of the vacant set containing a given vertex x. This algorithm is similar to the one used in the Bernoulli percolation case, but has some important modifications due to the dependence in our model.

We start the proof by reducing the complexity of the problem. We set, as in Section 4, $L = n^\gamma$, $\ell = \ln^2 n$, with $\gamma \in (0,1)$. Due to Lemma 4.1 it is sufficient to show that Theorem 1.1 holds with with P^{un} replaced by $Q^{un} = \mathsf{P} \circ \mathcal{X}^{-1}|_{D([0,un],V)}$.

Since we are looking for an upper bound on the vacant set, we can disregard the bridges Z^i in the concatenation $\mathcal{X}$ (cf. (4.2)). More precisely, we set $m = \lfloor un/(L+\ell) \rfloor$, and we observe that P-a.s. the vacant set

$$\mathcal{V}_n^u = V \setminus \big\{\mathcal{X}_t\big((Y^i),(Z^i)\big)\colon t \in [0,un]\big\} \tag{5.1}$$

is a subset of the *vacant set left by segments*, $\bar{\mathcal{V}}^u$,

$$\bar{\mathcal{V}}^u := V \bigg\backslash \bigcup_{i<m} \operatorname{Ran} Y^i, \quad \text{where } \operatorname{Ran} Y^i = Y^i_{[0,L]}. \tag{5.2}$$

Let $\bar{\mathcal{C}}_{\max} = \bar{\mathcal{C}}^u_{\max}$ and $\bar{\mathcal{C}}_x = \bar{\mathcal{C}}^u_x$ be the largest connected component, and the component containing x of $\bar{\mathcal{V}} = \bar{\mathcal{V}}^u$, respectively. Then the inequality

$$\mathsf{P}\big[|\bar{\mathcal{C}}_{\max}| \geq K \log n\big] \leq n\mathsf{P}\big[|\bar{\mathcal{C}}_x| \geq K \log n\big] \tag{5.3}$$

implies that to prove Theorem 1.1 it is sufficient to show the following proposition.

Proposition 5.1. *Let G be a connected graph on n vertices satisfying the assumptions* (A0), (A1), (A2) *and let $u > u_\star$. Then for every $\sigma > 0$ there exist $1 \leq c, K < \infty$ not depending on n (but depending on d, u, α_1, α_2) such that*

$$\mathsf{P}\big[\big|\bar{\mathcal{C}}^u_x\big| \geq K \operatorname{ld} n\big] \leq cn^{-\sigma-1}. \tag{5.4}$$

Proof. We prove this proposition by analysing the following BFS algorithm. During the run of the algorithm, all vertices in V are in one of four states: explored-vacant, explored-occupied, not-explored or in-queue. The set of vertices with the state in-queue is organised as a queue $\mathcal{Q}$, that is it is ordered, the vertices are added to its end and removed from its beginning. The vertices in $\mathcal{Q}$ wait to get explored.

Further, the state of any index $i \in \{0,\dots,m-1\}$ of a segment Y^i can be either free or tied. Note here that these states do not change the behaviour of the algorithm, but will be used for its analysis. Their meaning will be easier to understand as we get to Lemma 5.2.

When the algorithm starts, all vertices different from x are in not-explored state, x is in-queue, $\mathcal{Q} = (x)$, and all indices $i \in \{0,\dots,m-1\}$ are free.

At the step k of the algorithm, the first vertex y of the queue $\mathcal{Q}$ is removed from $\mathcal{Q}$. If $y \notin \bar{\mathcal{V}}$, then the state of y is changed to explored-occupied, and all indices i of segments intersecting y (i.e. $\{i\colon y \in Y^i\}$) become tied. On the other hand, if $y \in \bar{\mathcal{V}}$, then the state of y changes to explored-vacant and all non-explored neighbours of y in G are placed at the end of $\mathcal{Q}$, in other words, their state changes to in-queue. To avoid ambiguity, we suppose that V is equipped with an ordering and the neighbours of y are added to $\mathcal{Q}$ according to this ordering.

The algorithm stops if the queue $\mathcal{Q}$ is empty, or if the set of explored-vacant vertices has more than $K \operatorname{ld} n$ vertices. Since this set is subset of $\bar{\mathcal{C}}_x$ by construction, we know that only in the second case we have $|\bar{\mathcal{C}}_x| \geq K \operatorname{ld} n$. Hence, in order to establish Proposition 5.1, one only needs to show that

$$\text{there are } K, c > 0 \text{ such that, with probability at least } 1 - cn^{-\sigma-1}\text{, the algorithm finishes because the queue gets empty.} \tag{5.5}$$

To analyse the algorithm we need more notation. Throughout this section we fix

$$r = (7 \operatorname{ld} \operatorname{ld} n) \vee 2. \tag{5.6}$$

We let EV_k ($\mathsf{EO}_k, \mathsf{IQ}_k$) stand for the (random) set of vertices in the explored-vacant (explored-occupied, in-queue, respectively) state before the beginning of the kth step of the algorithm. Similarly, $\mathsf{F}_k, \mathsf{T}_k$ denote the sets of free and tied indices at this moment. We set $\mathsf{E}_k = \mathsf{EO}_k \cup \mathsf{EV}_k$ and let y_k be the vertex being explored in the kth step. In particular $y_1 = x$, $\mathsf{EV}_1 = \mathsf{EO}_1 = \mathsf{T}_1 = \varnothing$. Let $k_{\max}$ be the step when the algorithm finishes,

$$k_{\max} = \min\bigl\{k\colon\ |\mathsf{IQ}_k| = 0 \text{ or } |\mathsf{EV}_k| \geq K \operatorname{ld} n\bigr\}. \tag{5.7}$$

Observe that by construction $\mathsf{EO}_k \subset \partial_{\mathrm{e}}\mathsf{EV}_k$ and thus $\mathsf{E}_k \subset (\mathsf{EV}_k \cup \partial_{\mathrm{e}}\mathsf{EV}_k)$. Since G is d-regular, and since exactly one vertex is explored at every step, this implies

$$|\mathsf{E}_k| = k - 1 \leq dK \operatorname{ld} n \quad \text{for all } k \leq k_{\max}. \tag{5.8}$$

We further define a filtration $(\mathcal{A}_k)_{k\geq 1}$, where the σ-algebra $\mathcal{A}_k$ contains all information discovered by the algorithm before the kth step, that is

$$\mathcal{A}_k = \sigma(\mathsf{EV}_j, \mathsf{EO}_j, \mathsf{IQ}_j, \mathsf{F}_j, \mathsf{T}_j\colon\ j \leq k \wedge k_{\max}). \tag{5.9}$$

Observe that, due to the ordering that we use while adding vertices to $\mathcal{Q}$, the random variables $y_1, \dots, y_{k-1}$ are $\sigma(\mathsf{EO}_k, \mathsf{EV}_k)$-measurable.

To prove Proposition 5.1 we analyse the process recording the length of the queue, $q_k = |\mathsf{IQ}_k|$, $1 \leq k \leq k_{\max}$. We use $r_k = q_{k+1} - q_k$, $1 \leq k < k_{\max}$, to denote the size of its jumps. Since the graph G is d-regular, in step k, at most $d-1$ (d if $k = 1$) vertices are added to $\mathcal{Q}$, and every time exactly one vertex is removed from it. Therefore, $r_1 \in \{-1, d-1\}$ and $r_k \in \{-1, \dots, d-2\}$, for $k \geq 2$.

Roughly speaking, to prove (5.5) we will show that the process $(q_k)_{k\leq k_{\max}}$ has a 'down-drift.' For this, we need a lower bound on the probability that $r_k = -1$ given the past $\mathcal{A}_k$ of the algorithm. Since $r_k = -1$, whenever $y_k \notin \bar{\mathcal{V}}^u$, we have, on the event $\{k < k_{\max}\}$,

$$\begin{aligned}\mathsf{P}[r_k = -1|\mathcal{A}_k] &\geq \mathsf{P}[y_k \notin \bar{\mathcal{V}}|\mathcal{A}_k] = \mathsf{P}\biggl[y_k \in \bigcup_{i<m} \operatorname{Ran} Y^i \biggm| \mathcal{A}_k\biggr] \\ &\geq \mathsf{P}\biggl[y_k \in \bigcup_{i\in\mathsf{F}_k} \operatorname{Ran} Y^i \biggm| \mathcal{A}_k\biggr].\end{aligned} \tag{5.10}$$

The reason why we have made the distinction between the free and tied indices is made clear in the short lemma below.

Lemma 5.2. *Let $k \geq 1$ and $k_* = k \wedge k_{\max}$. Then, conditioned on the σ-algebra $\mathcal{A}_{k_*}$, the collection $(Y^i)_{i\in\mathsf{F}_{k_*}}$ is i.i.d. with marginal distribution $P^L[\cdot|H_{\mathsf{E}_{k_*}} > L]$.*

Proof. Let $\mathsf{H} = (\mathsf{EV}_j, \mathsf{EO}_j, \mathsf{IQ}_j, \mathsf{F}_j, \mathsf{T}_j)_{j\leq k_*}$ be the whole history of the algorithm until the time k_*. We use $\bar{H} = (V_j, O_j, Q_j, F_j, T_j)_{j\leq \bar{k}}$ to denote possible outcomes of H, here $\bar{k}$ is a positive integer. Since the Y^i's have marginal distribution P^L, it suffices to prove that for any $\bar{H}$ such that $\mathsf{P}[\mathsf{H} = \bar{H}] > 0$ and any measurable subsets A_i of $D([0, L], V)$,

$$\mathsf{P}\biggl[\bigcap_{i\in F_{\bar{k}}} \bigl\{Y^i \in A_i\bigr\}, \mathsf{H} = \bar{H}\biggr] = \biggl(\prod_{i\in F_{\bar{k}}} \mathsf{P}\bigl[Y^i \in A_i \,|\, \operatorname{Ran} Y^i \cap E_{\bar{k}} = \varnothing\bigr]\biggr)\mathsf{P}[\mathsf{H} = \bar{H}], \tag{5.11}$$

where $E_j = O_j \cup V_j$. Let us now analyse the event $\mathsf{H} = \bar{H}$ in detail. Let $\bar{y}_j = E_{j+1} \setminus E_j$ be the vertex explored in the jth step in the history $\bar{H}$. If this vertex is vacant, that is $\bar{y}_j \in V_{j+1} \setminus V_j$, then we know that $\bar{y}_j \notin \bigcap_{i<m} \operatorname{Ran} Y^i$. On the

other hand, if it is occupied, that is $\bar{y}_j \in O_{j+1} \setminus O_j$, then necessarily $\bar{y}_j \notin \bigcap_{i\in F_{j+1}} \operatorname{Ran} Y^i$, $\bar{y}_j \in \bigcup_{i\in T_{j+1}} \operatorname{Ran} Y^i$ and $\bar{y}_j \in \bigcap_{i\in T_{j+1}\setminus T_j} \operatorname{Ran} Y^i$. Therefore, the event $\mathsf{H} = \bar{H}$ can be written as

$$\bigcap_{j<\bar{k}:V_{j+1}\setminus V_j\neq\varnothing} \Big(\bar{y}_j \notin \bigcap_{i<m} \operatorname{Ran} Y^i\Big)$$
$$\cap \bigcap_{j<\bar{k}:O_{j+1}\setminus O_j\neq\varnothing} \Big\{\Big(\bar{y}_j \notin \bigcap_{i\in F_{j+1}} \operatorname{Ran} Y^i\Big) \cap \Big(\bar{y}_j \in \bigcup_{i\in T_{j+1}} \operatorname{Ran} Y^i\Big) \cap \Big(\bar{y}_j \in \bigcap_{i\in T_{j+1}\setminus T_j} \operatorname{Ran} Y^i\Big)\Big\}. \tag{5.12}$$

Collecting the events containing Y^i with $i \in F_{\bar{k}}$, using $F_j \supset F_{\bar{k}}$ for all $j \le \bar{k}$, this can be rearranged as

$$\bigcap_{i\in F_{\bar{k}}} \big(\operatorname{Ran} Y^i \cap E_{\bar{k}} = \varnothing\big) \cap f\big(\bar{H}, \big(\operatorname{Ran} Y^i\colon i \in T_{\bar{k}}\big)\big), \tag{5.13}$$

where f is some event depending only on $\bar{H}$ and $\operatorname{Ran} Y^i$ with $i \in T_{\bar{k}}$. Inserting this expression for $\mathsf{H} = \bar{H}$ into (5.11) and using the independence of Y^i's under P, the lemma follows. □

Lemma 5.2 implies that, on the event $\{k < k_{\max}\}$,

$$\mathsf{P}[r_k = -1|\mathcal{A}_k] \ge 1 - \mathsf{P}\Big[y_k \notin \bigcap_{i\in\mathsf{F}_k} \operatorname{Ran} Y^i \Big| \mathcal{A}_k\Big]$$
$$= 1 - \big(P[H_{\mathsf{E}_k\cup\{y_k\}} > L|H_{\mathsf{E}_k} > L]\big)^{|\mathsf{F}_k|}. \tag{5.14}$$

To bound (5.14), we will use Proposition 3.6 with $A = \mathsf{E}_k$, $y = y_k \in \partial_{\mathrm{e}}\mathsf{E}_k$ and $s = r$, see below (5.5). We first check its assumptions: Inequality (5.8) implies that for $n \ge c_K$, $B(\mathsf{E}_k, r) \le |\mathsf{E}_k|(d-1)^r \le \sqrt{n}$; $k \ge 2$ implies $\mathsf{E}_k \neq \varnothing$ and E_k is connected by construction. For $k \ge 2$, let $\mathcal{P}_k = \{\mathcal{F}_{\mathsf{E}_k}(y_k, r)$ is proper$\}$, see (3.28).

Take $\varepsilon_u > 0$ such that $u_\star(1+\varepsilon_u)^2 < u$. Since $r = 7\,\mathrm{ld}\,\mathrm{ld}\,n$ and $L = n^\gamma$, the error term in (3.29) is smaller than $c_K n^{\gamma-1}/\ln n$ which is much smaller than the leading term. Hence, for $n \ge c_{u,\alpha_2,K}$, on the event $\mathcal{P}_k$, we have by Proposition 3.6 that on $\{k < k_{\max}\}$,

$$P[H_{\mathsf{E}_k\cup y_k} > L|H_{\mathsf{E}_k} > L] \le \exp\Big\{\frac{-n^{\gamma-1}(d-2)^2}{d(d-1)(1+\varepsilon_u)}\Big\}. \tag{5.15}$$

We further define $\mathcal{G}_k = \{|\mathsf{F}_k| \ge u_\star n^{1-\gamma}(1+\varepsilon_u)^2\}$. Observe that both $\mathcal{P}_k$ and $\mathcal{G}_k$ are $\mathcal{A}_k$-measurable. Inserting (5.15) into (5.14) and using the definition (1.2) of $u_\star$, we get

$$\mathsf{P}[r_k = -1|\mathcal{A}_k] \ge \Big[1 - \exp\Big\{\frac{-n^{\gamma-1}(d-2)^2}{d(d-1)(1+\varepsilon_u)}\Big\}^{u_\star n^{1-\gamma}(1+\varepsilon_u)^2}\Big]\mathbf{1}_{\mathcal{G}_k\cap\mathcal{P}_k}$$
$$\ge (1+\delta_u)\frac{d-2}{d-1}\mathbf{1}_{\mathcal{G}_k\cap\mathcal{P}_k} \tag{5.16}$$

for some $\delta_u > 0$ on $\{k < k_{\max}\}$, provided $n \ge c_{u,\gamma}$.

To proceed we need to control the occurrence of $(\mathcal{G}_k \cap \mathcal{P}_k)^c$. Observe that, for $k \le k_{\max}$, $\mathcal{G}_k \supset \mathcal{G}_{k_{\max}}$.

Lemma 5.3. *For every $\sigma > 0$ there exists $c = c_{\sigma,u,K,\gamma}$ such that $\mathsf{P}[\mathcal{G}^c_{k_{\max}}] \le cn^{-\sigma-1}$.*

Proof. We set $M_n = \max_{z\in V} |\{i < m\colon Y^i \ni z\}|$ and we define the event $\bar{\mathcal{G}} = \{M_n \le \mathrm{ld}^2\, n\}$. Observe that on $\bar{\mathcal{G}}$ we have by (5.8)

$$|\mathsf{F}_k| = m - |\mathsf{T}_k| \ge m - |\mathsf{EO}_k|\,\mathrm{ld}^2\, n \ge m - dK\,\mathrm{ld}^3\, n, \tag{5.17}$$

which is larger than $u_\star n^{1-\gamma}(1+\varepsilon_u)^2$ for $n \geq c_{K,u}$. Therefore, $\mathcal{G}^c_{k_{\max}} \subset \bar{\mathcal{G}}^c$ for $n \geq c_{K,u}$.

It remains to bound $\mathsf{P}[\bar{\mathcal{G}}^c]$. First, note that,

$$\begin{aligned}\mathsf{P}\big[z \in \operatorname{Ran} Y^i\big] &= P[H_z \leq L] \leq P\big[N_L \geq n^{2\gamma}\big] + P\left[\bigcup_{j=1}^{\lfloor n^{2\gamma}\rfloor} \big\{Y^i_{\tau_j} = z\big\}\right] \\ &\overset{(4.10)}{\leq} c_\gamma \mathrm{e}^{-c' n^\gamma} + 2n^{1-\gamma} \leq c_\gamma n^{\gamma-1}. \end{aligned} \tag{5.18}$$

Using the bound above and an exponential Chebyshev-type inequality, we obtain that $\mathsf{P}[|\{i < m\colon Y^i \ni z\}| \geq \mathrm{ld}^2 n] \leq c_{u,\gamma}\mathrm{e}^{-\mathrm{ld}^2 n}$. Summing over z we get $\mathsf{P}[\bar{\mathcal{G}}^c_\zeta] \leq c_{u,\sigma,\gamma} n^{-\sigma-1}$ and the lemma follows. □

We further control the number of steps for which $\mathcal{P}_k$ does not hold. This is the content of the following proposition whose proof is postponed to the end of the section.

Proposition 5.4. *There are at most $cr K^2$ steps of the algorithm for which $\mathcal{P}^c_k$ occurs.*

To show (5.5) we now couple the process q with another process $(q'_k)_{k\geq 1}$ which is a random walk with drift such that $r'_k = q'_{k+1} - q'_k \in \{-1, d-2\}$ and $\mathsf{P}[r'_k = -1] = (1+\delta_u)\frac{d-2}{d-1}$ (see (5.16)). This implies that $\mathsf{E}[r'_k] < \delta'_u$ for a constant $\delta'_u < 0$.

The coupling is constructed so that q' can be used as an upper bound for q. This is done as follows. Let $k \in \{2, \dots, k_{\max}-1\}$. On $(\mathcal{G}_k \cap \mathcal{P}_k)^c$ we take r'_k independent of q. On $\mathcal{G}_k \cap \mathcal{P}_k$ we require that $r'_k = d-2$ whenever $r_k \geq 0$. This is possible because $\mathsf{P}[r_k \geq 0] \leq \mathsf{P}[r'_k = d-2]$ on $\mathcal{G}_k \cap \mathcal{P}_k$, due to (5.16). For $k = 1$ and $k \geq k_{\max}$, r'_k is independent of q.

As initial condition we take $q'_1 = crK^2(d-2) + d$. Intuitively speaking, this gives q' a security zone, for the steps in which $\mathcal{P}_k$ does not hold. With this setting, the Lemma 5.3 and Proposition 5.4 imply that $q'_k \geq q_k$ for all $k \leq k_{\max}$ with probability larger or equal to $1 - cn^{-\sigma-1}$.

We can finally show (5.5). The probability that the algorithm finishes due to $|\mathsf{EV}_{k_{\max}}| \geq K \,\mathrm{ld}\, n$ is bounded from above by

$$\mathsf{P}\Big[\min_{k<K\,\mathrm{ld}\,n} q_k > 0\Big] \leq cn^{-1-\sigma} + \mathsf{P}\Big[\min_{k<K\,\mathrm{ld}\,n} q'_k > 0\Big] \leq cn^{-1-\sigma} + \mathsf{P}\big[q'_{K\,\mathrm{ld}\,n} > 0\big] \leq c' n^{-1-\sigma}$$

for K, c' large enough, by an easy large deviation estimate for the random walk q', which has a negative drift. This finishes the proof of Proposition 5.1 and consequently of Theorem 1.1. □

It remains to show Proposition 5.4. The next lemma is the key step in its proof. It controls the tree excess of small (non-necessarily ball-like) sets.

Lemma 5.5. *Let $G = (V, \mathcal{E})$ with $|V| = n$ satisfy (A0) and (A1). Then for all $\kappa \geq 1$ and all connected sets $A \subset V$ such that $|A| \leq \kappa \,\mathrm{ld}\, n$*

$$\mathtt{tx}(A) \leq c\kappa^2 =: \alpha(\kappa). \tag{5.19}$$

Proof. Let $G_A = (A, \mathcal{E}_A)$ be the subgraph of G induced by A and let $s = \alpha_1 \,\mathrm{ld}\, n$. We call a cycle in G_A *short* if it has no more than $2s$ edges, otherwise we call it *long*.

Roughly speaking, the strategy to prove the lemma will be to erase edges belonging to short cycles, then to bound the amount of edges that could be still removed after that.

Fix a short cycle C and let $O_C = \{y \in A\colon C \subset B_{G_A}(y, s)\}$. Since A is connected, either $O_C = A$ or $|O_C| \geq s$. Further, since (A1) holds for G, it holds also for G_A. Therefore, if C, C' are two distinct short cycles, then O_C and $O_{C'}$ are disjoint. This implies that if $O_C = A$ for a short cycle C, then there is only one short cycle in A and that in any case, there are at most $|A|/s = \kappa/\alpha_1$ short cycles in G_A. From every of these disjoint short cycles we can erase one edge and G_A remains connected. Hence we erase at most κ/α_1 edges in this step.

After this removal, we obtain a graph $G'_A = (A, \mathcal{E}')$ with girth larger than $2s$. Recall from (2.3) that $\mathtt{tx}(A) = |\mathcal{E}_A| - |A| + 1$. Hence, since G_A and G'_A are both connected and G'_A was obtained by removing no more than $1 + \kappa/\alpha_1$ edges of G_A,

$$\mathtt{tx}(A) \le 1 + \kappa/\alpha_1 + \mathtt{tx}\big(G'_A\big). \tag{5.20}$$

To estimate the last term on the right-hand side, consider the set $D = \{(x, y) \in A^2\colon \operatorname{dist}_{G'_A}(x, y) \ge s\}$. Let $\gamma = (x_0, x_1, \dots, x_m, x_0)$ be a (necessarily long) cycle in G'_A. By removing the edge $\{x_0, x_m\}$, the size of the set D increases at least by $(\frac{s}{2} - 1)^2$. Indeed, before removing this edge any pair (x_i, x_j), for $0 \le i \le \frac{s}{2} - 1$, $0 \le m - j \le \frac{s}{2} - 1$ was not in D. However, after removing $\{x_0, x_m\}$, such a pair must be in D, since otherwise there would be a path in G'_A connecting x_i and x_j, not passing through the edge $\{x_0, x_m\}$ and having length at most s, thus there would be a short cycle in G'_A which is not possible. Since the size of D is at most $|A|^2$, it is not possible to remove more than $|A|^2/(\frac{s}{2} - 1)^2$ edges from G'_A while keeping it connected. Hence,

$$\mathtt{tx}\big(G'_A\big) \le \frac{|A|^2}{(s/2 - 1)^2}, \quad \text{which, for } n \ge c, \text{ is smaller or equal to } 16\kappa^2/{\alpha_1}^2. \tag{5.21}$$

The claims (5.20) and (5.21) imply that $\mathtt{tx}(A) \le 1 + (\kappa/\alpha_1) + (16\kappa^2/{\alpha_1}^2) \le c\kappa^2$, for $n \ge c'$. Lemma 5.5 now follows by possibly adjusting the constants. □

Proof of Proposition 5.4. The algorithm defined in the beginning of the proof of Proposition 5.1 induces a natural random tree structure $T = (\mathsf{E}_{k_{\max}}, \mathcal{E}_T)$. Namely, $\{y, z\} \in \mathcal{E}_T$ if and only if z was added to the queue during the exploration of y or vice-versa. By (5.8) we have $|\mathsf{E}_{k_{\max}}| \le dK \operatorname{ld} n =: \kappa \operatorname{ld} n$.

We now finish the proof of Proposition 5.4 in three lemmas which respectively control the number of y_k's for which (i), (iii) or (ii) of (3.28) do not hold. It is worth to remark that the arguments in these lemmas are purely deterministic and do not depend on the fact that T results from the previous BFS algorithm.

We start dealing with the condition (i) of (3.28). Recall the definition of $\alpha(\kappa)$ in Lemma 5.5 and that $r = (7 \operatorname{ld}\operatorname{ld} n) \vee 2$.

Lemma 5.6. *Let $B = \{y_k\colon k < k_{\max}$ and $\mathtt{tx}(\mathcal{F}_{\mathsf{E}_k}(y_k, r)) \ne 0\}$. Then, for large enough c, $|B| \le 2r\alpha(2\kappa)$ for all $n \ge c$.*

Proof. For $i < k_{\max}$, we define inductively a sequence γ_i of paths in V as follows. If $y_i \notin B$ or if y_i belongs to $\bigcup_{j<i} \operatorname{Ran}\gamma_j$, then $\gamma_i = \varnothing$. If $y_i \in B$ and $y_i \notin \bigcup_{j<i} \operatorname{Ran}\gamma_j$, then there is a cycle in $\mathcal{F}_{\mathsf{E}_i}(y_i, r)$ by definition of B and this cycle is unique by assumption (A1) (note that $\mathtt{tx}(\mathcal{F}_{\mathsf{E}_i}(y_i, r)) \le 1$ since $r \le \alpha_1 \operatorname{ld} n$ for $n \ge c$). In this case, we define γ_i as the unique path in $\mathcal{F}_{\mathsf{E}_i}(y_i, r)$ from y_i to this cycle, concatenated with the self-avoiding path exploring the whole cycle in one of the two directions.

Set $A = \{1 \le i < k_{\max}\colon \gamma_i \ne \varnothing\}$ and observe that $|\operatorname{Ran}\gamma_i| \le 2r$, for all $i \in A$, and $B \subset \bigcup_{j\in A} \operatorname{Ran}\gamma_j$, hence $|B| \le 2r|A|$.

It remains to show that $|A| \le \alpha(2\kappa)$. Assume the opposite. Let A_0 be any subset of A with $\alpha(2\kappa) + 1$ elements, set $R = \mathsf{E}_{k_{\max}} \cup \bigcup_{i\in A_0} \operatorname{Ran}\gamma_i$. Obviously,

$$|R| \le |\mathsf{E}_{k_{\max}}| + 2r\big(\alpha(2\kappa) + 1\big) \le 2\kappa \operatorname{ld} n \quad \text{for } n \ge c. \tag{5.22}$$

We claim that

$$\mathtt{tx}(R) \ge \alpha(2\kappa) + 1, \tag{5.23}$$

which together with (5.22) contradicts Lemma 5.5 and hence proves Lemma 5.6. The estimate (5.23) will follow if we can show that for all $i \in A_0$, we have $\mathtt{tx}(R_{i-1}) < \mathtt{tx}(R_i)$, where

$$R_0 = \varnothing \quad \text{and} \quad R_i = \{y_1, \dots, y_i\} \cup \bigcup_{j\in A_0, j\le i} \operatorname{Ran}\gamma_j, \quad 1 \le i \le |\mathsf{E}_{k_{\max}}|. \tag{5.24}$$

If $R_{i-1} \cap \operatorname{Ran} \gamma_i = \varnothing$, then this last claim is immediate, because γ_i then contains an additional cycle disjoint from R_{i-1}. Suppose now that $R_{i-1} \cap \operatorname{Ran} \gamma_i \neq \varnothing$. We will now find a cycle in R_i using an edge that is not already present in the graph induced by R_{i-1}. This will again imply that $\mathtt{tx}(R_{i-1}) < \mathtt{tx}(R_i)$, because by removing such an edge the graph remains connected and still has the graph induced by R_{i-1} as a subgraph. To find the cycle, note that $y_i \notin R_{i-1}$, because i can be in A_0 only if $\gamma_i \neq \varnothing$, which by construction can only happen if $\operatorname{Ran} \gamma_j \cap \{y_i\} = \varnothing$ for all $j < i$. Let $\bar{y}_i$ be the parent of y_i in the tree T. Then by construction of T, $\bar{y}_i \in R_{i-1}$. We now exhibit a cycle in R_i as follows: we start at $\bar{y}_i$ and connect $\bar{y}_i$ to y_i. We then follow the path γ_i from y_i to the first vertex x_0 belonging to R_{i-1}. Since R_{i-1} is connected, we can close our cycle by concatenating our path with a non-intersecting path from x_0 to $\bar{y}_i$ using only vertices in R_{i-1} and therefore not intersecting the previously constructed path from $\bar{y}_i$ to x_0. We have thus found a cycle in R_i using the edge $\{\bar{y}_i, y_i\}$. Since $y_i \notin R_{i-1}$, this edge is not present in R_{i-1} and it again follows that $\mathtt{tx}(R_{i-1}) < \mathtt{tx}(R_i)$. We have therefore proved (5.23) and thereby completed the proof of Lemma 5.6. □

We now treat condition (iii) of (3.28).

Lemma 5.7. *Let $B = \{y_k\colon$ there is path in $B(\mathsf{E}_k, r) \setminus \mathsf{E}_k$ from y_k to $\mathsf{E}_k \setminus \bar{y}_k\}$, where $\bar{y}_k$ is the parent of y_k in T. Then, for $n \geq c$, $|B| \leq 2r\alpha(2\kappa)$.*

Proof. The proof is analogous to the previous one. We define a sequence γ_i of paths in V as follows: If $y_i \in B$ and $y_i \notin \bigcup_{j<i} \operatorname{Ran} \gamma_j$, then let γ_i be a self-avoiding path connecting y_i to $\mathsf{E}_i = \{y_1, \dots, y_{i-1}\}$, whose first vertex after y_i is in $\mathcal{F}_{\mathsf{E}_i}(y_i, r)$ and whose length is at most $2r$. Provided n is large, such a path exists for any $y_i \in B$, because $\mathtt{tx}(B(y_i, r)) \leq 1$ (cf. (A1), (5.6)). Otherwise, we set $\gamma_i = \varnothing$. Defining A, A_0, R and R_i as in the proof of Lemma 5.6, we can again prove (5.22) and (5.23). The argument is the same as the one used below (5.23), except that we now only have to consider the case $R_{i-1} \cap \operatorname{Ran} \gamma_i \neq \varnothing$. □

Finally, we treat condition (ii) of (3.28).

Lemma 5.8. *With $B = \{y_k\colon$ y_k is neighbour in G of two vertices in $\mathsf{E}_k\}$, $|B| \leq \alpha(2\kappa)$.*

Proof. In this case we can remove the edges between y_i and E_i which are not in $\mathcal{E}_T$ from the subgraph of G induced by $\mathsf{E}_{k_{\max}}$ while keeping it connected. Since $|\mathsf{E}_{k_{\max}}| \leq \kappa \operatorname{ld} n$, Lemma 5.5 implies the result. □

Proposition 5.4 follows easily from last three lemmas. □

6. Super-critical regime

In this section we prove Theorem 1.2 stating the existence of a giant component for u smaller than $u_\star$. Since the proof is rather lengthy we first briefly outline its strategy. The strategy is inspired by the methods used for Bernoulli percolation. It has two major parts: First, we consider a modification of the piecewise independent measure and for such modification, we prove the existence of a sufficient amount of mesoscopic clusters even under a slightly increased value $u_n > u$. Second, by decreasing u_n back to the original value u, we prove that these clusters are connected by sprinkling. Both these parts are however rather non-trivial, due to the presence of the dependence.

To construct the mesoscopic clusters, we first show in Section 6.2 that the vacant set left by segments (see (5.2) and (6.14) below) on G locally resembles the vacant set of random interlacement on the d-regular tree $\mathbb{T}_d = (\mathbb{V}_d, \mathbb{E}_d)$ (Proposition 6.3). The behaviour of the random interlacement on $\mathbb{T}_d$ is well known [26] and its clusters can be controlled in terms of a particular branching process. This branching process will be super-critical for u's considered in this section.

The control by the branching process allows us to construct a sufficient amount of mesoscopic clusters for the vacant set left by segments. Since we are looking for a lower bound on the vacant set, we however cannot ignore the bridges as in the previous section. In Section 6.3 we show that the mesoscopic clusters of the vacant set left by segments are robust and the addition of the bridges to the picture typically does not destroy them, see Proposition 6.6.

Finally, in Section 6.4 we use a sprinkling well adapted to our model to prove Theorem 1.2. As discussed in the Introduction, in this sprinkling we erase randomly some segments, possibly in the middle of the trajectory. This can

possibly disconnect the trajectory. Therefore, to be able to extract a nearest-neighbour path in the end, we must add many additional bridges to the picture; the robustness proven in Section 6.3 must take them in consideration.

6.1. *Preliminaries*

We establish first the following technical consequence of assumption (A1) which will be needed later in this section.

Lemma 6.1. *Let $G = (V, \mathcal{E})$ be a graph satisfying assumptions* (A0), (A1). *Set $R = \lfloor \alpha_1 \operatorname{ld} n \rfloor$ to be the radius of* (A1) *and $r = R - \Delta$, for some $\Delta \in \{1, 2, \dots\}$. Then*

$$\big|\big\{x\colon \mathtt{tx}\big(B(x,r)\big) = 0\big\}\big| \geq \big(1 - (d-1)^{-\Delta}\big)|V|. \tag{6.1}$$

Proof. Let us consider the sets

$$A = \big\{x\colon \mathtt{tx}\big(B(x,r)\big) = 1\big\} \quad \text{and} \quad \tilde{A} = \big\{x\colon \mathtt{tx}\big(B(x,R)\big) = 1\big\}. \tag{6.2}$$

We first study the structure of the graph G restricted to $\tilde{A}$. For every point $x \in \tilde{A}$, there is exactly one cycle in $B(x, R)$. This cycle should contain at most $2R + 1$ vertices, otherwise it cannot be contained there. If $C \subset V$ is such a cycle, we define $\ell_C = \operatorname{diam} C = \lfloor |C|/2 \rfloor$ and $N_C(s) = \{x\colon C \subset B(x,s)\}$ for $s \leq R$. It is easy to see that (A1) implies $N_C(R) = B(C, R - \ell_C)$.

We now prove the following claim: the subgraph of G induced by $N_C(R)$ is composed by the cycle C with disjoint trees rooted at its vertices with depth $R - \ell_C$. Indeed, the graphs attached to every $y \in C$ should be trees because otherwise (A1) cannot hold. To see that they must be disjoint, suppose that they are not, that is there are two points $y, z \in C$ such that y and z are connected in $N_C(R) \setminus C$. This connection must be shorter than $2(R - \ell_C) + 1$. Joining this connection with the shortest connection of y and z in C, which is shorter than ℓ_C, we obtain a cycle different from C of length at most $2R - \ell_C + 1$, which is contained in $B(y, R)$. This, however, contradicts (A1).

The claim proved in the above paragraph implies that

$$\big|N_C(R)\big| = |C| + (d-2)|C|\big(1 + \cdots + (d-1)^{R-\ell_C-1}\big) = |C|(d-1)^{R-\ell_C}. \tag{6.3}$$

Since $N_C(r)$ is either empty or has a similar structure as $N_C(R)$

$$\big|N_C(R) \setminus N_C(r)\big| \geq |C|(d-1)^{r-\ell_C}\big((d-1)^{\Delta} - 1\big) \geq \big|N_C(r)\big|\big((d-1)^{\Delta} - 1\big). \tag{6.4}$$

By our assumptions, the set $\tilde{A}$ can be written as a disjoint union $\tilde{A} = \bigcup_{i=1}^{M} N_{C_i}(R)$ for some $M \in \mathbb{N}$ and cycles $C_1, \dots, C_M$. Similarly $A = \bigcup_{i \in U} N_{C_i}(r)$ for some $U \subset \{1, \dots, M\}$ which contains indices of cycles shorter than $2r + 1$. Therefore, using (6.3) and (6.4),

$$\begin{aligned} |V| \geq |\tilde{A}| &\geq \sum_{i \in U} \big|N_{C_i}(R)\big| = \sum_{i \in U} \big|N_{C_i}(r)\big| + \big|N_{C_i}(R) \setminus N_{C_i}(r)\big| \\ &\geq \sum_{i \in U} \big|N_{C_i}(r)\big|(d-1)^{\Delta} = (d-1)^{\Delta}|A|. \end{aligned} \tag{6.5}$$

Hence $|A| \leq |V|/(d-1)^{\Delta}$ and thus $|A^c| = |\{x\colon \mathtt{tx}(B(x,r)) = 0\}| \geq (1 - (d-1)^{-\Delta})|V|$. □

We now collect some notation used in the proof of Theorem 1.2. In what follows, we write $\mathbb{V} = \mathbb{V}_d$ for the set of vertices of the tree $\mathbb{T}_d$ and denote by o its root. In order to describe the clusters of random interlacement on the tree $\mathbb{T}_d$ we define the function $f : \mathbb{V} \to \mathbb{R}$ as

$$f(z) = \begin{cases} \frac{(d-2)^2}{d(d-1)}, & \text{if } z \neq o, \\ \frac{d-2}{d-1}, & \text{for } z = o. \end{cases} \tag{6.6}$$

We let $\mathbb{Q}_u^\star$ stand for the law on $\{0,1\}^{\mathbb{V}}$ which associates to the vertices $z \in \mathbb{V}$ independent Bernoulli random variables with success probability $\mathrm{e}^{-uf(z)}$. The following result of [26] provides the connection between this Bernoulli percolation and random interlacement on $\mathbb{T}_d$. This result will not be used in this paper, but is quoted here in order to provide the natural interpretation of the model that we have just introduced.

Theorem 6.2 ([26], Theorem 5.1 and (5.7)). *The connected component $\mathcal{C}_o \subseteq \mathbb{V}$ containing the root o has the same law under $\mathbb{Q}_u$ characterised by* (2.13) *as under $\mathbb{Q}_u^\star$.*

Note that, under the law $\mathbb{Q}_u^\star$ the cluster $\mathcal{C}_o \subseteq \mathbb{V}$ can be regarded as a branching process, where the ancestor in generation 0 is born with probability $\exp\{-u\frac{d-2}{d-1}\}$, and with binomial offspring distribution with parameters $d-1$ and p_u, where

$$p_u = \exp\left\{-\frac{u(d-2)^2}{d(d-1)}\right\}. \tag{6.7}$$

In order to deal with this branching process it is useful to define the expected number of offsprings as well as its logarithm in base $d-1$:

$$m_u = (d-1)p_u \quad \text{and} \quad v_u = \mathrm{ld}\, m_u = 1 - \frac{u(d-2)^2}{d(d-1)\ln(d-1)} = 1 - \frac{u}{u_\star}. \tag{6.8}$$

Observe that for $u < u_\star$, we have $m_u > 1 + c_u$ and $v_u \in (c_u, 1)$, for a constant $c_u > 0$.

For $u \in (0, u_\star)$, it will be convenient to fix a small $\varepsilon = \varepsilon(u) > 0$ such that the slightly increased intensity $u(1+\varepsilon)$ satisfies

$$u(1+\varepsilon) < \frac{u+u_\star}{2} \quad \text{and} \quad \frac{1}{4} + \frac{11}{4}\varepsilon < \frac{u_\star}{2(u+u_\star)}, \tag{6.9}$$

which by (6.8) implies that

$$\frac{3}{2}v_{u(1+\varepsilon)} - \frac{5}{4}v_{u(1-\varepsilon)} = \frac{1}{4} - \frac{u}{u_\star}\left(\frac{1}{4} + \frac{11}{4}\varepsilon\right) > 0. \tag{6.10}$$

Finally, we define

$$\beta = \frac{\alpha_1}{100} < \frac{1}{100} \quad \text{and} \quad \gamma = \gamma(u) = \frac{v_{u(1+\varepsilon)}\beta}{2} < \frac{\beta}{2}. \tag{6.11}$$

When $u \geq u_\star$ (which we allow in Section 6.2) we only require a weaker condition

$$0 < \gamma < \beta = \alpha_1/100 \leq 1/100. \tag{6.12}$$

We recall from Section 4 the segments Y^i with length $L = n^\gamma$ constructed on the probability space (Ω, P). We define

$$M_u = \lceil un/(L+\ell) \rceil \quad \text{for } u > 0. \tag{6.13}$$

We consider the *vacant set left by segments* $\bar{\mathcal{V}}^u = V \setminus \bigcup_{0 \leq i < M_u} \mathrm{Ran}\, Y^i$, and the corresponding random configuration $\xi_u \in \{0,1\}^V$ defined by

$$\xi_u = 1_{\bar{\mathcal{V}}^u} = 1\left\{V \setminus \bigcup_{0 \leq i < M_u} \mathrm{Ran}\, Y^i\right\}. \tag{6.14}$$

Given a configuration $\eta \in \{0,1\}^V$ or $\eta \in \{0,1\}^{\mathbb{V}}$, let

$$\mathcal{C}_x(\eta) \text{ be the connected component of } \mathrm{supp}\,\eta := \{y \in V\colon \eta(y) = 1\} \text{ containing the vertex } x, \tag{6.15}$$

and $\mathcal{C}_{\max}(\eta)$ be the largest such component.

For any fixed vertex $y \in V$, we define

$$B_y = B(y, \beta \operatorname{ld} n), \qquad B'_y = B(y, 5\beta \operatorname{ld} n) \subseteq V. \tag{6.16}$$

We also set

$$\mathbb{B} = B(o, \beta \operatorname{ld} n), \qquad \mathbb{B}' = B(o, 5\beta \operatorname{ld} n) \subset \mathbb{V}. \tag{6.17}$$

If y has a tree-like neighbourhood of radius $5\beta \operatorname{ld} n$,

$$\text{there is a graph isomorphism } \phi : B'_y \to \mathbb{B}' \text{ such that } \phi(y) = o. \tag{6.18}$$

In order to make the formulas less complicated, we will mostly identify the vertices of G and of $\mathbb{T}_d$ linked by this isomorphism and omit ϕ from the notation. The vertex y is always given by the context. In particular, for $z \in B_y$ we define $f(z) = f(\phi(z))$.

6.2. *Approximation by random interlacements*

With all notation in place, we can now approach the proof of Theorem 1.2. In this section we show that, provided $\mathtt{tx}(B'_y) = 0$, the component of the set $\bar{\mathcal{V}}^u \cap B_y = \operatorname{supp} \xi_u \cap B_y$ containing the centre y of B_y can, up to a small error, be controlled from above and from below by the branching process introduced in (6.6) and below. Note that for the next proposition it is not necessary to assume that $u < u_\star$.

Proposition 6.3. *Assume* (A0) *and* (A2), *and suppose that* $\mathtt{tx}(B'_y) = 0$. *Then for any* $u \geq 0$, $\varepsilon \in (0,1)$, *we can construct random sets* $\mathcal{C}^{u(1+\varepsilon)}$ *and* $\mathcal{C}^{u(1-\varepsilon)} \subseteq \mathbb{V}$ *distributed as* $\mathcal{C}_o$ *under* $\mathbb{Q}^\star_{u(1+\varepsilon)}$ *and* $\mathbb{Q}^\star_{u(1-\varepsilon)}$ *such that*

$$\mathsf{P}\Big[\mathcal{C}^{u(1+\varepsilon)} \cap \mathbb{B} \subseteq \mathcal{C}_y(1_{B_y} \cdot \xi_{\hat{u}}) \subseteq \mathcal{C}^{u(1-\varepsilon)} \text{ for all } \hat{u} \in \Big(u\Big(1 - \frac{\varepsilon}{2}\Big), u\Big(1 + \frac{\varepsilon}{2}\Big)\Big)\Big] \geq 1 - c_{\gamma,u,\varepsilon} n^{-2\beta}.$$

Remark 6.4. *Proposition* 6.3 *can also be interpreted as a control of the component of* $\bar{\mathcal{V}}^u \cap B_y$ *by random interlacement on* $\mathbb{T}_d$. *Indeed, due to Theorem* 6.2, *the sets* $\mathcal{C}^{u(1\pm\varepsilon)}$ *have the same distribution under the Bernoulli measure* $\mathbb{Q}^\star_{u(1\pm\varepsilon)}$ *as under the random interlacement measure* $\mathbb{Q}_{u(1\pm\varepsilon)}$.

Proof of Proposition 6.3. Throughout this proof, we write B, B' rather than B_y, B'_y. Our strategy resembles the proof of Theorem 5.1 in [26]. We first poissonise the number of trajectories entering in the definition of the configurations $\xi_{\hat{u}}$ for $\hat{u} \in (u(1-\varepsilon/2), u(1+\varepsilon/2))$, see (6.14). To this end we introduce two independent Poisson random variables Π_+ and Π_- defined on (Ω, P) with parameters $un^{1-\gamma}(1 + \frac{3\varepsilon}{4})$ and $un^{1-\gamma}(1 - \frac{3\varepsilon}{4})$, independent of all previously introduced random variables. We are going to compare $\xi_{\hat{u}}$ with the configurations ξ_-, ξ_+ defined by

$$\xi_- = 1\Big\{V \setminus \bigcup_{i<\Pi_-} \operatorname{Ran} Y^i\Big\} \quad \text{and} \quad \xi_+ = 1\Big\{V \setminus \bigcup_{i<\Pi_+} \operatorname{Ran} Y^i\Big\}. \tag{6.19}$$

Clearly, by a large deviation argument, since $M_u = un^{1-\gamma}(1 + \mathrm{o}(1))$,

$$\begin{aligned} &\mathsf{P}\big[\xi_+ \leq \xi_{\hat{u}} \leq \xi_- \text{ for all } \hat{u} \in \big(u(1-\varepsilon/2), u(1+\varepsilon/2)\big)\big] \\ &\quad \geq \mathsf{P}[\Pi_- \leq M_{u(1-\varepsilon/2)} < M_{u(1+\varepsilon/2)} \leq \Pi_+] \geq 1 - c_{\gamma,u,\varepsilon} \exp\big\{-cn^{1-\gamma}\big\}. \end{aligned} \tag{6.20}$$

Next, we will dominate $\xi_\pm$ from above and from below by a collection of i.i.d. Bernoulli random variables. For every $z \in B$, we define the set D_z as the set of descendants of z in B, that is

$$D_z = \big\{z' \in B\colon \text{ any path from } z' \text{ to } y \text{ meets either } z \text{ or some vertex in } B^c\big\},$$

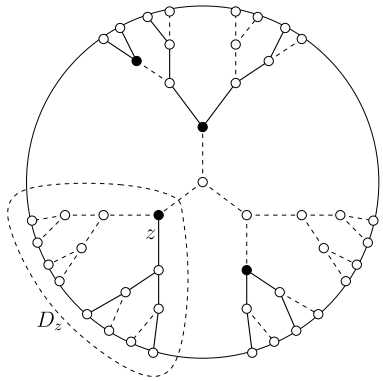

Fig. 2. The trace left by some pieces in B and the set D_z. The black and white circles represent respectively the values 0 and 1 for the variables $\tilde{\xi}_\pm(z)$ defined in (6.22).

see Fig. 2. Consider the following disjoint subsets of $D([0,L],V)$, cf. [26], (5.3),

$$\begin{aligned} W_z &= \big\{Y \in D\big([0,L],V\big)\colon z \in \operatorname{Ran} Y \cap B \subset D_z\big\}, \quad z \in B, \\ W &= \big\{Y \in D\big([0,L],V\big)\colon \operatorname{Ran} Y \cap B \neq \varnothing\big\} \Big\backslash \bigcup_{z\in B} W_z. \end{aligned} \tag{6.21}$$

In particular, all trajectories in W must enter B, then exit B' and enter B again, see Fig. 2. We define the random configurations $\tilde{\xi}_+, \tilde{\xi}_- \in \{0,1\}^B$ on (Ω, P) by

$$\begin{aligned} \tilde{\xi}_-(z) &= 1\big\{Y^i \notin W_z \ \forall i < \Pi_-\big\}, \quad z \in B, \\ \tilde{\xi}_+(z) &= 1\big\{Y^i \notin W_z \ \forall i < \Pi_+\big\}, \quad z \in B, \end{aligned} \tag{6.22}$$

see Fig. 2 again. Since the sets W_z are disjoint for distinct z's, the variables $\xi_+(z)$ will be independent for distinct z's due to the Poissonian character of Π_+ (the same will also hold for $\xi_-(x)$), see Lemma 6.5 below. We further consider the random variable

$$\mathcal{Z} = 1\big\{Y^i \notin W \ \forall i < \Pi_+\big\}. \tag{6.23}$$

Observe that

$$\text{on the event } \{\mathcal{Z}=0\}, \mathcal{C}_y(1_B \cdot \xi_-) = \mathcal{C}_y(\tilde{\xi}_-) \text{ and } \mathcal{C}_y(1_B \cdot \xi_+) = \mathcal{C}_y(\tilde{\xi}_+). \tag{6.24}$$

The following lemma shows that the laws of $\tilde{\xi}_-$ and $\tilde{\xi}_+$ on $\{0,1\}^B$ are comparable with the laws $\mathbb{Q}^\star_{u(1-3\varepsilon/4)}$ and $\mathbb{Q}^\star_{u(1+3\varepsilon/4)}$ of Bernoulli percolation introduced above, restricted to $\{0,1\}^{\mathbb{B}}$ (which by assumption can be identified with $\{0,1\}^B$).

Lemma 6.5. *For $\pm$ denoting either $+$ or $-$, the events $(\{\tilde{\xi}_\pm(z)=1\})_{z\in B}$ are independent and satisfy*

$$\begin{aligned} &\big|\mathsf{P}\big[\tilde{\xi}_+(z)=1\big] - \mathrm{e}^{-u(1+3\varepsilon/4)f(z)}\big| \le c_{\gamma,u} n^{-\gamma/3} \quad \textit{and} \\ &\big|\mathsf{P}\big[\tilde{\xi}_-(z)=1\big] - \mathrm{e}^{-u(1-3\varepsilon/4)f(z)}\big| \le c_{\gamma,u} n^{-\gamma/3}. \end{aligned} \tag{6.25}$$

Before we prove this lemma, we complete the proof of Proposition 6.3. For $z \in \mathbb{V}$ let

$$\begin{aligned} p_z^+ &= \mathrm{e}^{-u(1+\varepsilon)f(z)}, \qquad q_z^+ = \mathsf{P}\big[\tilde{\xi}_+\big(\phi^{-1}(z)\big)=1\big], \\ p_z^- &= \mathrm{e}^{-u(1-\varepsilon)f(z)}, \qquad q_z^- = \mathsf{P}\big[\tilde{\xi}_-\big(\phi^{-1}(z)\big)=1\big]. \end{aligned} \tag{6.26}$$

Then Lemma 6.5 implies that for $n \geq c'_{\gamma,u,\varepsilon}$,

$$p_z^+ \leq q_z^+ \leq q_z^- \leq p_z^- \quad \text{for all } z \in \mathbb{B}. \tag{6.27}$$

We now construct the sets $\mathcal{C}^{u(1\pm\varepsilon)}$ as stated in the proposition by adding to our probability space (Ω, P) a collection $\{U_z^+, U_z^-\}_{z\in\mathbb{V}}$ of independent Bernoulli-distributed random variables which will fine tune the values $q_z^\pm$ to match the $p_z^\pm$'s:

- For every $z \in \mathbb{V} \setminus \mathbb{B}$, the parameters of U_z^+ and U_z^- are p_z^+ and p_z^-.
- For $z \in \mathbb{B}$, U_z^+ and U_z^- have parameters p_z^+/q_z^+ and $(p_z^- - q_z^-)/(1 - q_z^-)$.

We then define $\mathcal{C}^{u(1\pm\varepsilon)}$ by

$$\mathcal{C}^{u(1+\varepsilon)} = \mathcal{C}_o\big(\big((\tilde{\xi}_+(z) \wedge U_z^+)1_{z\in\mathbb{B}} + U_z^+ 1_{z\in\mathbb{V}\setminus\mathbb{B}}\big)_{z\in\mathbb{V}}\big), \tag{6.28}$$

$$\mathcal{C}^{u(1-\varepsilon)} = \mathcal{C}_o\big(\big((\tilde{\xi}_-(z) \vee U_z^-)1_{z\in\mathbb{B}} + U_z^- 1_{z\in\mathbb{V}\setminus\mathbb{B}}\big)_{z\in\mathbb{V}}\big). \tag{6.29}$$

Note that we then have

$$\mathcal{C}^{u(1+\varepsilon)} \subset \mathcal{C}_y(\tilde{\xi}_+) \subseteq \mathcal{C}_y(\tilde{\xi}_-) \subset \mathcal{C}^{u(1-\varepsilon)}. \tag{6.30}$$

Since the variables $\tilde{\xi}_\pm$ and $U_z^\pm$ are all independent (cf. Lemma 6.5), it is elementary to check that the laws of $\mathcal{C}^{u(1\pm\varepsilon)}$ agree with those of $\mathcal{C}_o$ under $\mathbb{Q}^\star_{u(1\pm\varepsilon)}$ for large n. Moreover, we have by (6.24) that on the event $\{\mathcal{Z} = 0\} \cap \{\xi_+ \leq \xi_{\hat{u}} \leq \xi_-\}$,

$$\mathcal{C}^{u(1+\varepsilon)} \cap \mathbb{B} \subseteq \mathcal{C}_y(1_B \cdot \xi_+) \subseteq \mathcal{C}_y(1_B \cdot \xi_{\hat{u}}) \subseteq \mathcal{C}_y(1_B \cdot \xi_-) \subseteq \mathcal{C}^{u(1-\varepsilon)} \cap \mathbb{B}. \tag{6.31}$$

Since we already know the bound (6.20), it thus only remains to prove that

$$\mathsf{P}[\mathcal{Z} \neq 0] \leq c_u n^{-2\beta}. \tag{6.32}$$

If $\mathcal{Z} \neq 0$, there is an $i < \Pi_+$ such that $Y^i \in W$. Since $\mathtt{tx}(B') = 0$, if $Y^i \in W$, then there exist times $t_1 < t_2 < t_3$ such that $Y^i_{t_1} \in B$, $Y^i_{t_2} \notin B'$ and again $Y^i_{t_3} \in B$, see (6.21). Using the strong Markov property we thus get

$$\mathsf{P}\big[Y^i \in W\big] = P^L[W] \leq P^L[H_B < L] \sup_{w\in V\setminus B'} P^L_w[H_B < L]. \tag{6.33}$$

Note that by stationarity of the random walk with respect to the uniform distribution,

$$P^L[H_B < L] \leq E^L\left[\sum_{k=0}^{N_L} 1\{\hat{X}_k \in B\}\right] = E^L[N_L]|B|/n = L|B|/n. \tag{6.34}$$

Using Lemma 3.4 for the second term on the right-hand side of (6.33), we hence obtain

$$\mathsf{P}\big[Y^i \in W\big] \leq c_\gamma n^{\gamma+\beta-1}\big(cn^{\gamma-4\beta} + \mathrm{e}^{-c'n^\gamma}\big) \leq c_\gamma n^{2\gamma-3\beta-1}. \tag{6.35}$$

Since $\mathcal{Z}$ has Poisson distribution with parameter $un^{1-\gamma}(1 + 3\varepsilon/4)P^L[W]$, using $\gamma < \beta$,

$$\mathsf{P}[\mathcal{Z} \neq 0] = 1 - \exp\big\{-u(1 + 3\varepsilon/4)n^{1-\gamma} P^L[W]\big\} \leq c_u n^{-2\beta}. \tag{6.36}$$

Up to Lemma 6.5, this completes the proof of Proposition 6.3. □

Proof of Lemma 6.5. Since the sets W_z, $z \in B$, are mutually disjoint and Π_- is Poisson distributed, independent of the Y^i's, the random variables $|\{Y^i \in W_z\colon i < \Pi_-\}|$, $z \in B$, are independent Poisson random variables with parameters

$un^{1-\gamma}(1-3\varepsilon/4)\mathsf{P}[Y^0 \in W_z]$. In particular, since $\tilde{\xi}_-(z) = 1\{|\{Y^i \in W_z\colon i < \Pi_-\}| = 0\}$, the events $(\{\tilde{\xi}_-(z) = 1\})_{z\in B}$ are indeed independent. Moreover, we have

$$\mathsf{P}\big[\tilde{\xi}_-(z) = 1\big] = \mathsf{P}\big[\big|\{Y^i \in W_z\colon i < \Pi_-\}\big| = 0\big] = \mathrm{e}^{-un^{1-\gamma}(1-3\varepsilon/4)\mathsf{P}[Y^0\in W_z]}. \tag{6.37}$$

The above arguments apply also to $\tilde{\xi}_+$ and yield the analogous claims. Since the function e^{-x} is Lipschitz with constant 1 on $[0,\infty)$, we see that the left-hand sides of (6.25) are bounded by $c_u|n^{1-\gamma}P^L[W_z] - f(z)|$. It is therefore sufficient to prove that

$$\big|n^{1-\gamma}P^L[W_z] - f(z)\big| \le c_\gamma n^{-\gamma/3} \quad \text{for any } z \in B. \tag{6.38}$$

Note the relation of this approximation with (1.6).

Conditioning on the number of jumps N_L made by X in the time interval $[0, L]$ and using independence of N_L and the discrete skeleton $\hat{X}$, we have

$$P^L[W_z] = \sum_{r\ge 0} P[N_L = r]P\big[z \in \{\hat{X}_0, \dots, \hat{X}_r\} \cap B \subset D_z\big]. \tag{6.39}$$

Let us fix any r such that $\lfloor 2^{-1}n^\gamma\rfloor \le r \le \lfloor 2n^\gamma\rfloor$ and throughout the rest of this proof write $A_z = B \setminus D_z$. Summing over all possible times k when $\hat{X}$ first visits z and applying the simple Markov property, we obtain

$$P\big[z \in \{\hat{X}_0, \dots, \hat{X}_r\} \cap B \subset D_z\big] = \sum_{0\le k\le r} P[\hat{H}_{A_z\cup\{z\}} = k, \hat{X}_k = z]P_z[\hat{H}_{A_z} > r-k], \tag{6.40}$$

where we are using the convention that $\hat{H}_\varnothing = \infty$, which occurs in the last probability when $z = y$, in which case $A_y = B \setminus D_y = \varnothing$. Using reversibility of $\hat{X}$ with respect to the uniform distribution on the first probability in the product, we deduce that

$$P\big[z \in \{\hat{X}_0, \dots, \hat{X}_r\} \cap B \subset D_z\big] = \frac{1}{n}\sum_{0\le k\le r} P_z\big[\hat{H}^+_{A_z\cup\{z\}} > k\big]P_z[\hat{H}_{A_z} > r-k]. \tag{6.41}$$

We now claim that the following estimates hold uniformly for all $n^{\gamma/2} \le k \le r - n^{\gamma/2}$:

$$\sup_{z'\in V\setminus B'} P_{z'}[\hat{H}_B \le k] \le c_\gamma n^{-3\beta} \quad \text{and} \quad \sup_{z'\in B'} P_{z'}[\hat{H}_{V\setminus B'} \ge k] \le c_\gamma(\log n)n^{-\gamma/2}. \tag{6.42}$$

Indeed, the first estimate follows from Lemma 3.4 and the choice of $\gamma < \beta$ in (6.11), while the second estimate in (6.42) follows from the Chebyshev inequality and the bound $E_{z'}[\hat{H}_{V\setminus B'}] \le c_\beta \log n$, which is an elementary estimate on the expected amount of time it takes a one-dimensional biased random walk to reach the level $5\beta\,\mathrm{ld}\,n$.

For any k as above, it follows from (6.42) and the strong Markov property applied at time $\hat{H}_{V\setminus B'}$ that

$$\big|P_z\big[\hat{H}^+_{A_z\cup\{z\}} > k\big] - P_z\big[\hat{H}^+_{A_z\cup\{z\}} > \hat{H}_{V\setminus B'}\big]\big| \le c_\gamma\big(n^{-3\beta} + n^{-\gamma/3}\big). \tag{6.43}$$

We now relate the second probability on the left-hand side to the escape probability to infinity from $\mathbb{B}\cup\{z\}$ for the random walk on the tree $\mathbb{T}_d$. By the strong Markov property applied at time $\hat{H}_{V\setminus B'}$, we have (identifying z and A_z with corresponding objects on $\mathbb{T}_d$)

$$P_z^{\mathbb{T}_d}\big[\hat{H}^+_{A_z\cup\{z\}} = \infty\big] \le P_z\big[\hat{H}^+_{A_z\cup\{z\}} > \hat{H}_{V\setminus B'}\big] \le \frac{P_z^{\mathbb{T}_d}[\hat{H}^+_{A_z\cup\{z\}} = \infty]}{\inf_{z'\in\mathbb{V}\setminus\mathbb{B}'} P_{z'}^{\mathbb{T}_d}[H_{\mathbb{B}} = \infty]}. \tag{6.44}$$

By another elementary estimate on the biased random walk $(\mathrm{dist}(B, \hat{X}_n))_{n\ge 0}$, we have

$$\inf_{z'\in\mathbb{V}\setminus\mathbb{B}'} P_{z'}^{\mathbb{T}_d}[H_{\mathbb{B}} = \infty] \ge 1 - cn^{-4\beta}. \tag{6.45}$$

Collecting the above estimates we obtain that for any $n^{\gamma/2} \le k \le r - n^{\gamma/2}$,

$$\left|P_z\big[\hat{H}^+_{A_z\cup\{z\}} > k\big] - P_z^{\mathbb{T}_d}\big[\hat{H}^+_{A_z\cup\{z\}} = \infty\big]\right| \le c_\gamma n^{-\gamma/3}, \tag{6.46}$$

and the same computations with $\hat{H}^+_{A_z\cup\{z\}}$ replaced by $\hat{H}_{A_z}$ show that

$$\left|P_z[\hat{H}_{A_z} > r-k] - P_z^{\mathbb{T}_d}[\hat{H}_{A_z} = \infty]\right| \le c_\gamma n^{-\gamma/3}. \tag{6.47}$$

With estimates on one-dimensional random walk, we can compute the escape probabilities for random walk on the infinite tree explicitly. Indeed, by applying the simple Markov property at time 1, and then computing the probability that a nearest-neighbour biased random walk on the integers does not return to 0 when started at 1, we obtain

$$P_z^{\mathbb{T}_d}\big[\hat{H}^+_{A_z\cup\{z\}} = \infty\big] = \begin{cases} \frac{d-1}{d} \times \frac{d-2}{d-1} = \frac{d-2}{d}, & \text{if } z \ne y, \\ \frac{d-2}{d-1}, & \text{if } z = y, \end{cases} \tag{6.48}$$

and similarly, by the convention that $\hat{H}_\varnothing = \infty$,

$$P_z^{\mathbb{T}_d}[\hat{H}_{A_z} = \infty] = \begin{cases} \frac{d-2}{d-1}, & \text{if } z \ne y, \\ 1, & \text{if } z = y. \end{cases} \tag{6.49}$$

Note that in both cases, the product of the two probabilities just computed equals $f(z)$, cf. (6.6). Inserting the estimates (6.46) and (6.47) into (6.41), we therefore infer that for any r such that $2^{-1}n^\gamma \le r \le 2n^\gamma$,

$$\left|P\big[z \in \{\hat{X}_0, \dots, \hat{X}_r\} \cap B \subset D_z\big] - rn^{-1} f(z)\right| \le c_\gamma n^{-1+(2\gamma/3)}. \tag{6.50}$$

Using this estimate and the large deviation bound on N_L from (4.10) in (6.39), we obtain that

$$\left|P^L[W_z] - Ln^{-1} f(z)\right| \le c_\gamma n^{-1+(2\gamma/3)}, \tag{6.51}$$

hence (6.38). This completes the proof of Lemma 6.5 and thus of Proposition 6.3. □

6.3. *Existence of mesoscopic components*

We now use the results of the last section to establish the existence of many mesoscopic components in the (appropriately modified) vacant set. In order to state the precise result, we need, as we have discussed before, to introduce the long-range bridges that are necessary to perform the sprinkling.

Recall from Section 4 that $a_i = Y_0^i$ and $b_i = Y_L^i$ denote the start- and end-point of the segment Y^i, $i \ge 1$. On the same probability space (Ω, P), we now define a family of $D([0,\ell], V)$-valued random variables $Z^{i,j}$, $i \in \mathbb{N}$, $j \in \{1, \dots, \lfloor \ln n \rfloor\}$, with law characterized by the following:

$$\begin{aligned}&\text{conditionally on } a_i, b_i, \text{ the } Z^{i,j}\text{'s are independent, independent of the } Y^i\text{'s,}\\ &\text{and have distribution } P^\ell_{b_i, a_{i+j}}.\end{aligned} \tag{6.52}$$

We call $Z^{i,j}$'s the *long-range bridges*. Given the Y^i's and $Z^{i,j}$'s as above, we denote by $\xi'_u \in \{0,1\}^V$ the indicator function of the vacant set left by them, i.e.

$$\xi'_u = 1\Big\{V \setminus \bigcup_{i<M_u,\, j\le \ln n} \big\{\operatorname{Ran} Y^i \cup \operatorname{Ran} Z^{i,j}\big\}\Big\}. \tag{6.53}$$

From definitions of ξ_u and ξ'_u it follows that $\xi'_u \le \xi_u$.

We now show that the configuration ξ'_u has many mesoscopic components. More precisely, the following proposition shows that with high probability, a constant proportion of vertices is contained in components of ξ'_u with size of order $n^{v_u(1+\varepsilon)^\beta}$.

Proposition 6.6. *For $0 < u < u_\star$, there exist constants c_1, c_2 depending on α_1, α_2 and u, such that*

$$\mathsf{P}\big[\big|\big\{x \in V\colon \big|\mathcal{C}_x(\xi'_u)\big| \geq c_1 n^{v_{u(1+\varepsilon)}\beta}\big\}\big| \geq c_1 n\big] \geq 1 - c_2 \exp\big\{-c \ln^3 n\big\}. \tag{6.54}$$

The proof of the proposition has two parts. First, in Lemma 6.8, we establish a similar result for the configuration ξ_u defined in (6.14) as the indicator of the complement of the segments. We then show that many of them survive adding the long-range bridges which will prove the Proposition 6.6.

6.3.1. *Robust mesoscopic components for ξ_u*
In order to ensure that adding the long-range bridges does not destroy the components of ξ_u of size $c_1 n^{v_{u(1+\varepsilon)}\beta}$, we should make them more robust. We therefore impose the following more restrictive conditions on the components to be found.

Definition 6.7. *Let η be a configuration in $\{0,1\}^V$ and set for $l \in \mathbb{N}$*

$$\mathcal{C}^l_x(\eta) = \big\{y \in \partial_{\mathrm{i}} B(x,l)\colon \ y \text{ is connected to } x \text{ by a path in } \mathcal{C}_x(\eta) \cap B(x,l)\big\}. \tag{6.55}$$

(*Note that $\mathcal{C}^l_x(\eta)$ is contained in, but not necessarily equal to $\mathcal{C}_x(\eta) \cap \partial_{\mathrm{i}} B(x,l)$.*) *Given a positive parameter h, a given site $x \in V$ is said to be h-proper under the configuration η, if $\mathtt{tx}(B(x, 3\beta \operatorname{ld} n)) = 0$ and the following two conditions hold* (*recall* (6.8)):

$$\begin{aligned}
&\text{(i)}\ \ |\mathcal{C}^{\beta \operatorname{ld} n}_x(\eta)| \geq h m^{\beta \operatorname{ld} n}_{u(1+\varepsilon)} = h n^{v_{u(1+\varepsilon)}\beta},\\
&\text{(ii)}\ \ |\mathcal{C}^l_y(\eta)| \leq m^{(5/4)l}_{u(1-\varepsilon)} \text{ for all } y \in B(x, \beta \operatorname{ld} n),\ l \in [l_0, l_1] \cap \mathbb{N}, \text{ where } l_0 = \lceil 10 \ln\ln n / v_{u(1-\varepsilon)} \rceil,\\
&\qquad l_1 = \lfloor \beta \operatorname{ld} n \rfloor.
\end{aligned} \tag{6.56}$$

The next lemma proves the existence of many proper sites.

Lemma 6.8. *For β, ε as in* (6.11), (6.9) *there exist constants $c_3(u)$, $c_4(u)$ and $c(\alpha_1, u)$ such that*

$$\mathsf{P}\big[\big|\{x \in V\colon x \text{ is } c_3\text{-proper under } \xi_u\}\big| \geq c_4 n\big] \geq 1 - c \exp\big\{-\ln^3 n\big\}. \tag{6.57}$$

Proof. We first show with Proposition 6.3 and estimates on branching processes that the expected number of proper vertices is of order n and then we show that this number is concentrated around its expectation. Throughout this proof, we abbreviate $m_{u(1+\varepsilon)}$ and $m_{u(1-\varepsilon)}$ by m_+ and m_-.

Consider $x \in V$ such that $\mathtt{tx}(B(x, \alpha_1 \operatorname{ld} n)) = 0$. We estimate the probability that x satisfies condition (6.56)(i) with $h > 0$ to be chosen. Since $5\beta < \alpha_1$, using Proposition 6.3 with $u(1+\varepsilon) < (u + u_\star)/2$ (cf. (6.9)),

$$\mathsf{P}\big[\big|\mathcal{C}^{\beta \operatorname{ld} n}_x(\xi_u)\big| < h m^{\beta \operatorname{ld} n}_+\big] \leq \mathbb{Q}^\star_{u(1+\varepsilon)}\big[\big|\mathcal{C}^{\beta \operatorname{ld} n}_o\big| < h m^{\beta \operatorname{ld} n}_+\big] + c_u n^{-2\beta}. \tag{6.58}$$

For $u(1+\varepsilon) < (u + u_\star)/2$, the branching process induced by $\mathbb{Q}^\star_{u(1+\varepsilon)}$ is supercritical. Hence, [5], Theorem 2, p. 9, implies that for $h < c_u$ chosen small enough, the first term on the right-hand side of (6.58) is bounded by $1 - c_u$ for $n \geq c'_u$. Therefore, letting h be some strictly positive constant $c_3 < c_u$,

$$\sup_{n \geq c_u} \mathsf{P}\big[\big|\mathcal{C}^{\beta \operatorname{ld} n}_x(\xi_u)\big| < c_3 m^{\beta \operatorname{ld} n}_+\big] < 1 - c_u. \tag{6.59}$$

We now treat condition (6.56)(ii). Since $m_u > 1 + c_u$, we can use [4], Theorem 4, to find a $\theta_u \in (0, c_u)$ such that (here, $\mathbb{E}^\star_u$ denotes $\mathbb{Q}^\star_u$-expectation)

$$H_u = \sup_n \sup_{l \geq 0} \mathbb{E}^\star_u\left[\exp\left\{\theta_u \frac{|\mathcal{C}^l_y|}{m^l_u}\right\}\right] < \infty. \tag{6.60}$$

We claim that for l_0 and l_1 as in (6.56), and ε as in (6.9),

$$\mathbb{Q}^{\star}_{u(1-\varepsilon)}\big[\big|\mathcal{C}^l_o\big| > m_-^{(5/4)l} \text{ for some } l \geq l_0\big] \leq c_u \exp\big\{-c'_u \ln^2 n\big\}. \tag{6.61}$$

Indeed, by the exponential Chebyshev inequality, the left-hand side of (6.61) can be bounded from above by

$$\begin{aligned}
&\sum_{l=l_0}^{\infty} \mathbb{Q}^{\star}_{u(1-\varepsilon)}\left[\exp\left\{\theta_{u(1-\varepsilon)} \frac{|\mathcal{C}^l_o|}{m^l_-}\right\} > \exp\big\{\theta_u m_-^{(1/4)l}\big\}\right] \\
&\quad \overset{(6.60)}{\leq} \sum_{l=l_0}^{\infty} H_{u(1-\varepsilon)} \exp\big\{-\theta_{u(1-\varepsilon)} m_-^{(1/4)l}\big\} \\
&\quad \leq H_{u(1-\varepsilon)} \exp\left\{-\theta_{u(1-\varepsilon)} \exp\left\{\left\lceil \frac{10 \ln\ln n}{\mathrm{ld}\, m_-}\right\rceil \frac{1}{4} (\mathrm{ld}\, m_-)(\ln d)\right\}\right\} \sum_{k=0}^{\infty} \mathrm{e}^{-\theta_{u(1-\varepsilon)} m_-^{l_0} (m_-^{(1/4)k} - 1)} \\
&\quad \overset{(\ln d \geq 1)}{\leq} c_u \exp\big\{-c'_u \ln^2 n\big\}.
\end{aligned} \tag{6.62}$$

This proves (6.61). It follows that

$$\begin{aligned}
&\mathsf{P}\big[\big|\mathcal{C}^l_y(\xi_u)\big| > m_-^{(5/4)l} \text{ for some } y \in B(x, \beta\, \mathrm{ld}\, n), l_0 \leq l \leq l_1\big] \\
&\quad \overset{\text{Proposition 6.3}}{\leq} cn^{\beta} \big\{\mathbb{Q}^{\star}_{u(1-\varepsilon)}\big[\big|\mathcal{C}^l_o\big| > m_-^{(5/4)l} \text{ for some } l \geq l_0\big] + c_u n^{-2\beta}\big\} \overset{(6.61)}{\leq} c_u n^{-\beta}.
\end{aligned} \tag{6.63}$$

The above bound, together with (6.59), allows us to conclude that for all $n \geq c'_u$, $\mathsf{P}[x \text{ is } c_3\text{-proper under } \xi_u] \geq c_u > 0$. Summing this probability over the vertices x with $\mathtt{tx}(B(x, 3\beta\, \mathrm{ld}\, n)) = 0$ (which have positive proportion by Lemma 6.1) we obtain that

$$\mathsf{E}\big[\big|\{x \in V\colon x \text{ is } c_3\text{-proper under } \xi_u\}\big|\big] \geq c_u n. \tag{6.64}$$

We now show that the number of c_3-proper points concentrates around its expectation. To this end we use a concentration inequality in [16], Lemma 1.2. We first consider a slightly modified configuration $\check{\xi} \in \{0,1\}^V$, where we consider only the first $[2n^{\gamma}]$ jumps of each Y^i:

$$\check{\xi} = \check{\xi}\big(Y^0, Y^1, \dots, Y^{M_u - 1}\big) = 1\left\{V \setminus \bigcup_{i < M_u} Y^i_{[0, \tau_{[2n^{\gamma}]}(Y^i) \wedge L]}\right\}, \tag{6.65}$$

where $\tau_k(Y^i)$ is the time of the kth jump of Y^i (we set $\tau_k(Y_i) = L$ if Y^i jumps less than k-times). We define a function

$$f\big(Y^0, \dots, Y^{M_u - 1}\big) = \big|\{x \in V\colon x \text{ is } c_3\text{-proper under } \check{\xi}\}\big|. \tag{6.66}$$

We claim that, writing $\vec{Y}$ for $(Y^0, \dots, Y^{M_u - 1})$,

$$\text{if } \vec{Y} \text{ and } \vec{Y}' \text{ differ in at most one coordinate, then } |f(\vec{Y}) - f(\vec{Y}')| \leq cn^{\gamma + \beta}. \tag{6.67}$$

Indeed, changing one segment Y^i, we can change at most $2n^{\gamma}$ values of $\check{\xi}$. Moreover, the event that a given point $x \in V$ is c_3-proper under $\check{\xi}$ only depends on the values of $\check{\xi}$ in $B(x, \beta\, \mathrm{ld}\, n)$, which has volume bounded by cn^{β}. This gives (6.67). Note that

$$\begin{aligned}
\mathsf{E}[f] \;&\geq\; \mathsf{E}[f 1_{\xi_u = \check{\xi}}] = \mathsf{E}\big[\big|\{x \in V\colon x \text{ is } c_3\text{-proper under } \xi_u\}\big| \cdot 1_{\xi_u = \check{\xi}}\big] \\
&\overset{(6.64)}{\geq} c_u n - n \cdot \mathsf{P}[\xi_u \neq \check{\xi}] \quad \text{for } n \geq c_u.
\end{aligned} \tag{6.68}$$

The bound (4.10) implies that

$$\mathsf{P}[\xi_u \neq \check{\xi}] \leq M_u P\big[N_L \geq 2n^\gamma\big] \leq \exp\big\{-c_u n^\gamma\big\}. \tag{6.69}$$

Hence, we have that $\mathsf{E}[f] \geq c_u n$ for $n \geq c'_u$. Setting $t = \frac{1}{2}c_u n$, with the same constant as in the lower bound on $\mathsf{E}[f]$, we obtain

$$\begin{aligned}\mathsf{P}\big[\big|\{x \in V\colon \text{ is } c_3\text{-proper under } \xi_u\}\big| \leq t\big] &\leq \mathsf{P}[\xi_u \neq \check{\xi}] + \mathsf{P}\big[\mathsf{E}[f] - f \geq t\big] \\ &\leq \exp\big\{-c_u n^\gamma\big\} + 2\exp\big\{-c_u n^{2-1+\gamma-2(\gamma+\beta)}\big\},\end{aligned} \tag{6.70}$$

where we have used Lemma 1.2 in [16], together with (6.67), in the last inequality. Since $1 - \gamma - 2\beta \geq 1 - 3\beta > 0$, this estimate is more than enough to imply (6.57) for appropriately chosen constants c_4 and c. This concludes the proof of Lemma 6.8. □

6.3.2. *Robustness of proper sites*

In this subsection we prove that the components around h-proper sites (as in Definition 6.7) are really robust with respect to perturbation. Observe that the following lemma is completely deterministic.

Lemma 6.9. *Let β, ε be as in (6.9), (6.11) and define the class Ξ of configurations in $\{0,1\}^V$,*

$$\Xi = \big\{\eta \in \{0,1\}^V\colon \big|\{x \in V\colon x \text{ is } c_3\text{-proper}\}\big| \geq c_4 n\big\}. \tag{6.71}$$

Let $\eta \in \Xi$ and $\eta' \in \{0,1\}^V$ be such that $\eta'(z) \neq \eta(z)$ for at most $n^{1-\gamma}\ln^5 n$ vertices $z \in V$. Then there exists a constant $c(\alpha_1, u)$, such that

$$\big|\big\{x \in V\colon \big|\mathcal{C}_x(\eta')\big| \geq c n^{v_{u(1+\varepsilon)}\beta}\big\}\big| \geq cn. \tag{6.72}$$

Proof. In this proof, we use the word "proper" to mean "c_3-proper under η" and use m_+, m_-, v_+ and v_- to abbreviate $m_{u(1+\varepsilon)}$, $m_{u(1-\varepsilon)}$, $v_{u(1+\varepsilon)}$ and $v_{u(1-\varepsilon)}$. We will use the term *string* to refer to a self-avoiding path on V with length $l_1 = \lfloor \beta \,\mathrm{ld}\, n \rfloor$, as in (6.56). For $\eta \in \Xi$, we are going to choose a particular collection Γ_η of strings, which will be contained in $\operatorname{supp}\eta$, as follows. First, we take a collection of proper vertices $\Pi = \{x_1, \dots, x_{\lfloor c_4 n\rfloor}\} \subseteq V$, according to some pre-defined order. Again using some arbitrary order, for each $l \leq \lfloor c_4 n \rfloor$, we insert into Γ_η $\lfloor c_3 n^{v_+\beta}\rfloor$ distinct strings starting at $x_l \in \Pi$ and contained in $\operatorname{supp}\eta$. Such a collection exists due to (6.56)(i) (see also (6.55)). Denoting by $|\Gamma_\eta|$ the number of strings in Γ_η, we have

$$|\Gamma_\eta| = \lfloor c_4 n \rfloor \cdot \big\lfloor c_3 n^{v_+\beta}\big\rfloor. \tag{6.73}$$

Since for all $l \leq \lfloor c_4 n \rfloor$, $B(x_l, 2\beta \log n)$ has tree excess zero,

$$\text{every string in } \Gamma_\eta \text{ is uniquely determined by its end-points.} \tag{6.74}$$

Let S_y be the number of strings in Γ_η intersecting y. We claim that, for any given $y \in V$,

$$S_y \leq c_u (\log n)^{c'_u} \cdot n^{(5/4)v_-\beta}. \tag{6.75}$$

To show this claim, observe that the fact that the starting point of every string in Γ_η is proper together with (6.56)(ii) imply that if there is a string intersecting y, then

$$\big|\mathcal{C}_y^l(\eta)\big| \leq m_-^{(5/4)l} \quad \text{for every integer } l \in [l_0, l_1]. \tag{6.76}$$

We bound S_y by splitting the set of strings intersecting y in the following way:

$$S_y = \sum_{l=0}^{l_1} \#\{\text{strings in } \Gamma_\eta \text{ intersecting } y \text{ and starting at distance } l \text{ from } y\}. \tag{6.77}$$

Since the strings are contained in $\operatorname{supp}\eta$, using (6.74), for $n \geq c_u$,we obtain

$$S_y \leq \sum_{l=0}^{l_1} \big|\mathcal{C}_y^l(\eta)\big|\big|\mathcal{C}_y^{l_1-l}(\eta)\big| \leq 2\sum_{l=0}^{l_0-1} \big|\mathcal{C}_y^l(\eta)\big|\big|\mathcal{C}_y^{l_1-l}(\eta)\big| + 2\sum_{l=l_0}^{l_1/2} \big|\mathcal{C}_y^l(\eta)\big|\big|\mathcal{C}_y^{l_1-l}(\eta)\big|. \tag{6.78}$$

Using (6.76), the bound $|\mathcal{C}_y^k| \leq c(d-1)^k$ for $k < l_0$, and $l_0 \leq c_u \ln\ln n$, we get

$$S_y \leq c_u \ln\ln n \cdot (d-1)^{l_0} m_-^{(5/4)l_1} + 2\sum_{l=0}^{l_1/2} m_-^{(5/4)l} m_-^{(5/4)(l_1-l)}. \tag{6.79}$$

From (6.8), it follows that $m_-^{l_1} \leq n^{v_- \beta}$. Hence,

$$S_y \leq c_u (\ln n)^{c_u'} n^{(5/4)v_-\beta} + 2\sum_{l=0}^{l_1/2} n^{(5/4)v_-\beta} \leq c_u (\ln n)^{c_u'} \cdot n^{(5/4)v_-\beta}. \tag{6.80}$$

This proves (6.75).

Our next step is to show that there exists c_u such that

$$\text{for } n \geq c_u, \text{ at least half of the strings of } \Gamma_\eta \text{ are contained in } \operatorname{supp}\eta'. \tag{6.81}$$

Indeed, we know that $\eta'(z) \neq \eta(z)$ for at most $n^{1-\gamma}\ln^5 n$ vertices $z \in V$. This, together with (6.75), implies that at most $c_u(\ln n)^{c_u'} n^{1-\gamma+(5/4)v_-\beta}$ strings in Γ_η are not contained in $\operatorname{supp}\eta'$. Since $\gamma = v_+\beta/2$ (cf. (6.11)), we obtain by (6.10) that $1-\gamma+\frac{5}{4}v_-\beta < 1+v_+\beta$. Therefore, due to (6.73), for $n \geq c_u''$, at least half of the strings in Γ_η are contained in $\operatorname{supp}\eta'$. This gives us (6.81).

Let us recall that in the construction of the set Γ_η, we have chosen a collection Π of $\lfloor c_4 n\rfloor$ proper vertices in V, and for each of these vertices, we have picked $\lfloor c_3 n^{v_+\beta}\rfloor$ strings starting at x_l. We claim that

$$\begin{aligned}&\text{for } n \geq c_u, \text{ at least } \lfloor \tfrac{c_4}{8} n\rfloor \text{ of the vertices in } \Pi \text{ have at least } \tfrac{c_3}{8} n^{v_+\beta}\\ &\text{of their strings contained in } \operatorname{supp}\eta'.\end{aligned} \tag{6.82}$$

Indeed, otherwise the number of strings in Γ_η contained in $\operatorname{supp}\eta'$ would be bounded by

$$\frac{c_4}{8} n \cdot c_3 n^{v_+\beta} + \frac{7c_4}{8} n \cdot \frac{c_3}{8} n^{v_+\beta} \leq \frac{c_3 c_4}{4} n^{1+v_+\beta}, \tag{6.83}$$

contradicting (6.81) for $n \geq c_u$.

Since distinct strings starting on a vertex in Π have distinct end points, each vertex x as in (6.82) satisfies $|\mathcal{C}_x(\eta')| \geq \frac{c_3}{8} n^{v_+\beta}$. Choosing c as $\frac{c_3 \wedge c_4}{8}$, we deduce (6.72) for $n \geq c_u$. By possibly decreasing $c = c_u$ in such a way that $\lfloor cn\rfloor = 0$ for the other finitely many values of n, we obtain Lemma 6.9, and thus complete the proof of Proposition 6.6. □

6.3.3. *Mesoscopic components for ξ_u'*

With Lemmas 6.8 and 6.9, we have all tools to finish the proof of Proposition 6.6 stating the existence of many mesoscopic components of the complement of the segments and the long-range bridges.

Proof of Proposition 6.6. The configurations ξ_u and ξ_u' differ only on vertices visited by the bridges. Hence, setting $\mathcal{D} = |\{z \in V\colon\ \xi_u(z) \neq \xi_u'(z)\}|$, denoting by $N^{i,j}$ the number of jumps of the bridge $Z^{i,j}$, for $n \geq \mathrm{e}^{u_\star}$, and using $M_u \leq c_u n^{1-\gamma}$, we obtain

$$\begin{aligned}\mathsf{P}\big[\mathcal{D} > n^{1-\gamma}\ln^5 n\big] &\leq \mathsf{P}\Bigg[\sum_{i<M_u,\, j\leq \ln n} N^{i,j} > c_u n^{1-\gamma}\ln^4 n\Bigg]\\ &\leq \mathsf{P}\Bigg[\bigcup_{i\leq M_u,\, j\leq \ln n} \big\{N^{i,j} > \ln^3 n\big\}\Bigg] \leq c_u \exp\big\{-c\ln^3 n\big\},\end{aligned} \tag{6.84}$$

where we have used (4.11) from Lemma 4.2 in the last inequality. By possibly increasing c_u we conclude that the equation above holds for every $n \geq 1$. Finally, taking c as in Lemma 6.9, using Lemma 6.9,

$$\begin{aligned}&\mathsf{P}\big[\big|\{x \in V\colon \big|\mathcal{C}_x(\xi'_u)\big| \geq cn^{v_u(1+\varepsilon)^\beta}\}\big| < cn\big]\\ &\quad \leq \mathsf{P}\big[\big|\{x\colon x \text{ is } c_3\text{-proper under } \xi_u\}\big| < c_4 n\big] + \mathsf{P}\big[\mathcal{D} > n^{1-\gamma}\ln^5 n\big].\end{aligned} \tag{6.85}$$

By Lemma 6.8 and (6.84) the last expression is bounded by $c_u \mathrm{e}^{-c\ln^3 n}$. The proof of Proposition 6.6 is then finished by choosing the constants c_1 and c_2 appropriately. □

6.4. *Proof of Theorem 1.2 and sprinkling.*

We can now approach the second part of the proof of Theorem 1.2, that is the sprinkling construction.

Proof of Theorem 1.2. Let u be as in the theorem and choose $\delta > 0$ such that

$$2\delta < 1 - \frac{1}{200} \overset{(6.11)}{<} 1 - \gamma(u). \tag{6.86}$$

Set u_n and u' as

$$u' = \frac{u + u_\star}{2}, \qquad u_n = \big(u + n^{-\delta}\big) \wedge u', \tag{6.87}$$

Throughout this proof, we write ε and γ for $\varepsilon = \varepsilon(u')$ and $\gamma = \gamma(u')$ as in (6.9) and (6.11) with u replaced by u'.

The strategy of this proof is the following: we apply Proposition 6.6 to $\xi'_{u'}$ as defined in (6.53). This will show that with high probability, there are at least $c_1 n$ vertices in components of volume $c_1 n^{v_{u'}(1+\varepsilon)^\beta}$ in $\operatorname{supp}\xi'_{u_n} \supseteq \operatorname{supp}\xi'_{u'}$. In what we call the sprinkling construction, we then erase some of the Y^i's in the definition of ξ'_{u_n}, and thereby increase the configuration ξ'_{u_n} to a new configuration ξ^{sp}. By construction, the sprinkled configuration ξ^{sp} will be close in distribution to the vacant set left by the random walk trajectory $X_{[0,un]}$. Moreover, we will prove that with high probability some of the components of $\operatorname{supp}\xi'_{u_n}$ will merge and form a component of size ρn as we increase ξ'_{u_n} to ξ^{sp}, thus proving Theorem 1.2.

We divide the proof into the following three steps: in the first step, we construct the sprinkled configuration ξ^{sp} and reduce Theorem 1.2 to an estimate on ξ^{sp}. In the second step, we apply Proposition 6.6 to prove that the original configuration ξ'_{u_n} is sufficiently well-behaved. In the third and final step, we deduce that with high probability, $\operatorname{supp}\xi^{\mathrm{sp}}$ has a component with volume at least ρn and conclude.

Step 1: The sprinkling construction. For the sprinkling construction, we use an auxiliary probability space $(\{0,1\}^{M_{u_n}}, \mathcal{Q})$, for M_{u_n} defined in (6.13). Under the measure $\mathcal{Q}$, the canonical coordinates $(R_k)_{0\leq k<M_{u_n}}$ are i.i.d. Bernoulli random variables with parameter

$$q = n^{-2\delta}. \tag{6.88}$$

Recall that the configuration ξ'_{u_n} was defined in (6.53) as the indicator function of the set of vertices not visited by Y^i and $Z^{i,j}$ for $i < M_{u_n}$ and $j \leq \lfloor \ln n \rfloor$, constructed on a suitable probability space (Ω, P). On the probability space $(\Omega \times \{0,1\}^{M_{u_n}}, \mathsf{P} \otimes \mathcal{Q})$, we will now construct from ξ'_{u_n} the sprinkled configuration ξ^{sp}, roughly according the following procedure: first, we remove all Y^i's such that $R_i = 1$. If possible, we then construct a trajectory by linking the remaining segments with the bridges $Z^{i,j}$, see Fig. 3 for a sketch.

For the precise construction, we define the increasing random sequence k_i of indices $k < M_{u_n}$ for which $R_k = 0$; the sub-indices i here run from 0 to the random variable

$$I = \big|\{k\colon 0 \leq k < M_{u_n}, R_k = 0\}\big|. \tag{6.89}$$

We now construct a function that concatenates the Y^i's with $R_i = 0$, $i < M_{u_n}$, and some of the bridges $Z^{i,j}$ into an element of $D([0, un], V)$, for u as in the theorem. There are two situations in which this construction fails. First, if

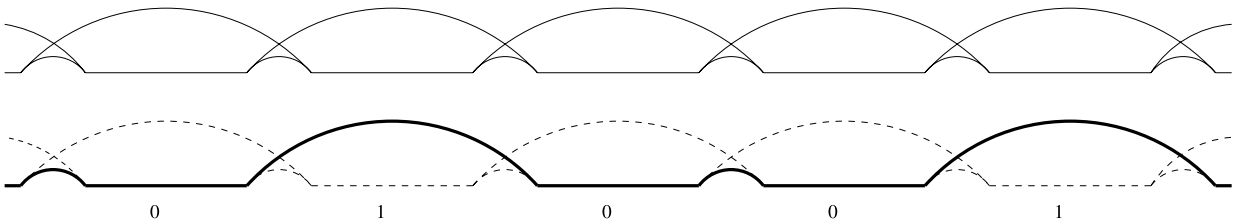

Fig. 3. Schematic illustration the choice performed by ψ. The horizontal lines are the trajectories of the segments Y^i and the arcs are the trajectories of the bridges $Z^{i,j}$. The top picture illustrates the configuration ξ'_{u_n}. The sprinkled configuration ξ^{sp} obtained from the realization of the Bernoulli random variables R_i associated to Y^i is on the bottom picture.

$I < M_u$, then there are not enough segments Y^i left. Second, if $k_{i+1} - k_i > \lfloor \ln n \rfloor$ for some $i \le I$, there is no bridge connecting Y^{k_i} to $Y^{k_{i+1}}$. Let us hence refer to the intersection of the complements of these events as the good event $\mathcal{G}$,

$$\mathcal{G} = \{I \ge M_u\} \cap \big\{k_{i+1} - k_i \le \lfloor \ln n \rfloor \text{ for all } 1 \le i \le M_u\big\} \subseteq \{0,1\}^{M_{u_n}}. \tag{6.90}$$

Letting ∂ be some arbitrary constant trajectory of length un, we now define $\psi : \Omega \times \{0,1\}^{M_{u_n}} \to D([0,un],V)$ as

$$\psi = \begin{cases} \partial, & \text{on } \Omega \times \mathcal{G}^c, \\ \mathcal{X}\big(Y^{k_1}, Z^{k_1,k_2-k_1}, Y^{k_2}, Z^{k_2,k_3-k_2}, \dots, Y^{k_{M_u}}, Z^{k_{M_u},k_{M_u+1}-k_{M_u}}\big)\big|_{D([0,un],V)}, & \text{otherwise,} \end{cases} \tag{6.91}$$

where $|_{D([0,un],V)}$ denotes the restriction to $D([0,un],V)$ and $\mathcal{X}$ is the concatenation mapping defined in (4.2). (Here we abuse the notation slightly. The mapping $\mathcal{X}$ takes infinite number of arguments, however since $un \le M_u(L+\ell)$ the restriction to $D([0,un],V)$ does not depend on the arguments which we do not specify.) The sprinkled configuration is then defined as the indicator function of the vacant set left by the concatenated trajectory,

$$\xi^{\mathrm{sp}} = 1\{V \setminus X_{[0,un]}\} \circ \psi, \tag{6.92}$$

where we have used the notation X for the canonical coordinate process on the space $D([0,un],V)$.

By construction, we then have $\operatorname{supp} \xi'_{u_n} \subseteq \operatorname{supp} \xi^{\mathrm{sp}}$ on $\Omega \times \mathcal{G}$. Moreover, conditionally on $\Omega \times \mathcal{G}$, the concatenation ψ is distributed according to the piecewise independent measure Q^{un} defined in Section 4.

Let us now see that the event $\mathcal{G}$ is indeed typical. The random variable I is binomially distributed with expectation $M_{u_n}(1-q) = M_u + n^{1-\gamma-\delta}(1+\mathrm{o}(1))$. With the help of a Chernoff bound (see Lemma 1.1 in [16]), we find that

$$\mathcal{Q}[I < M_u] \le \exp\big(-c_{u,\delta} n^{2(1-\gamma-\delta)}/M_{u_n}\big) \le \exp\big(-c_{u,\delta} n^{1-\gamma-2\delta}\big). \tag{6.93}$$

Together with a simple union bound, this implies that

$$\begin{aligned} \mathcal{Q}\big[\mathcal{G}^c\big] &\le \mathcal{Q}\big[\exists i \le I\colon k_{i+1} - k_i > \lfloor \ln n \rfloor\big] + \mathcal{Q}[I < M_u] \\ &\le c_u n^{1-\gamma} q^{\lfloor \ln n \rfloor} + \exp\big(-c_{u,\delta} n^{1-\gamma-2\delta}\big) \\ &\overset{(6.88)}{\le} c_u n^{1-\gamma} n^{-2\delta \lfloor \ln n \rfloor} + \exp\big(-c_{u,\delta} n^{1-\gamma-2\delta}\big) \le c_{u,\sigma,\delta} n^{-\sigma}. \end{aligned} \tag{6.94}$$

In particular, the distribution $\psi \circ (\mathsf{P} \otimes \mathcal{Q})$ is close to Q. Indeed, for any $\mathcal{F}_{un}$-measurable event A, we have $Q^{un}[A] = \mathsf{P} \otimes \mathcal{Q}[\psi^{-1}(A) | \Omega \times \mathcal{G}]$, and therefore using an easy calculation

$$\big|Q^{un}(A) - (\mathsf{P} \otimes \mathcal{Q})\big(\psi^{-1}(A)\big)\big| \le \mathcal{Q}\big(\mathcal{G}^c\big) \le c_{u,\sigma,\delta} n^{-\sigma}. \tag{6.95}$$

Thanks to this estimate, we know that the sprinkled configuration ξ^{sp} is close in distribution to the vacant set left by a trajectory under the piecewise independent measure Q^{un}. Together with Lemma 4.1, we can now reduce our task to proving the estimate in Theorem 1.2 for the configuration left by ξ^{sp}. We set

$$\rho = c_1/100, \tag{6.96}$$

where c_1 was defined in Proposition 6.6. By Lemma 4.1,

$$P\big[\big|\mathcal{C}^u_{\max}\big| < \rho n\big] \le Q^{un}\big[\big|\mathcal{C}_{\max}\big(1\{V \setminus X_{[0,un]}\}\big)\big| < \rho n\big] + \exp\big\{-c_u \ln^2 n\big\}, \tag{6.97}$$

which by (6.95) implies that

$$P\big[\big|\mathcal{C}^u_{\max}\big| < \rho n\big] \le \mathsf{P} \otimes \mathcal{Q}\big[\big|\mathcal{C}_{\max}\big(\xi^{\mathrm{sp}}\big)\big| < \rho n\big] + c_{u,\delta,\sigma} n^{-\sigma}. \tag{6.98}$$

It is therefore sufficient to show that

$$\mathsf{P} \otimes \mathcal{Q}\big[\big|\mathcal{C}_{\max}\big(\xi^{\mathrm{sp}}\big)\big| < \rho n\big] \le c_{u,\delta,\sigma} n^{-\sigma}. \tag{6.99}$$

Step 2: ξ'_{u_n} is well-behaved with high probability. In the second step, we apply previous estimates in order to deduce that ξ'_{u_n} has the properties we will use to show that a component of size ρn appears in $\operatorname{supp} \xi^{\mathrm{sp}}$.

Let $\mathcal{C}_1, \mathcal{C}_2, \dots$ be the connected components of $\operatorname{supp} \xi'_{u_n}$ ordered according to their volume, $\mathcal{C}_1$ being the largest component, and define the random variable $\kappa \in \mathbb{N} \cup \{\infty\}$ as the smallest integer such that

$$|\mathcal{C}_1| + \cdots + |\mathcal{C}_\kappa| \ge c_1 n = 100\rho n, \tag{6.100}$$

provided such an integer exists, and $\kappa = \infty$ otherwise. Note that if $\operatorname{supp} \xi'_{u_n}$ contains many large clusters (in the sense of Proposition 6.6), then κ is small. More precisely, we have the following event inclusion,

$$\big\{\big|\big\{x \in V\colon \big|\mathcal{C}_x\big(\xi'_{u_n}\big)\big| \ge c_1 n^{v_{u'(1+\varepsilon)^\beta}}\big\}\big| \ge c_1 n\big\} \subseteq \big\{\kappa \le \big\lceil n^{1-v_{u'(1+\varepsilon)^\beta}}\big\rceil\big\} =: \mathcal{K} \subset \Omega. \tag{6.101}$$

Hence, Proposition 6.6 (and monotonicity of $\xi'_{\cdot}$) is more than enough to imply that

$$\mathsf{P}\big[\mathcal{K}^c\big] \le c_{u,\sigma} n^{-\sigma}. \tag{6.102}$$

We further define an event $\mathcal{S} \subset \Omega$ as the event that the numbers of jumps N^i of all Y^i's and the total length $N^{i,j}$ of all $Z^{i,j}$'s appearing in the construction of ξ'_{u_n} do not exceed their expected value too significantly:

$$\mathcal{S} = \big\{N^i \le 2n^\gamma \ \forall i < M_{u_n}\big\} \cap \Bigg\{\sum_{i < M_{u_n}} \sum_{j \le \ln n} N^{i,j} \le \ln^5 n \cdot n^{1-\gamma}\Bigg\}. \tag{6.103}$$

By Lemma 4.2 we know that

$$\mathsf{P}\big[\mathcal{S}^c\big] \le c_{u,\sigma} n^{-\sigma}. \tag{6.104}$$

The estimates (6.102) and (6.104) will allow us to prove the required estimate (6.99) in the last step by considering only configurations ξ'_{u_n} satisfying the properties in $\mathcal{K} \cap \mathcal{S}$.

Step 3: ξ^{sp} has a large component with high probability. Finally, we prove the required estimate (6.99) by showing that the random deletion of Y^i's in the construction of ξ^{sp} does make a component of size ρn appear with high probability.

We define an event $\mathcal{P}$ as

$$\mathcal{P} = \left\{\begin{array}{l}\text{There exists a partition of } \{1,\dots,\kappa\} \text{ into sets } A \text{ and } B,\\ \text{such that } \mathcal{A} = \bigcup_{a \in A} \mathcal{C}_a \text{ is not connected to } \mathcal{B} = \bigcup_{b \in B} \mathcal{C}_b \text{ in}\\ \operatorname{supp} \xi^{\mathrm{sp}}, \text{ and both } |\mathcal{A}| \text{ and } |\mathcal{B}| \text{ are larger than } 10\rho n\end{array}\right\}. \tag{6.105}$$

In analogy with the proof of Proposition 3.1 in [3], we claim that

$$\big\{\big|\mathcal{C}_{\max}\big(\xi^{\mathrm{sp}}\big)\big| < \rho n\big\} \cap \{\kappa < \infty\} \subseteq \mathcal{P}. \tag{6.106}$$

To see this, we consider the equivalence relation $\sim$ on the set $\{1,\dots,\kappa\}$ given by $j \sim j'$ if and only if $\mathcal{C}_j$ is connected to $\mathcal{C}_{j'}$ in $\operatorname{supp} \xi^{\mathrm{sp}}$. Then every equivalence class corresponds to one component of $\operatorname{supp} \xi^{\mathrm{sp}}$. In particular, if all

components of $\operatorname{supp}\xi^{\mathrm{sp}}$ are smaller than ρn, then the sum of $|\mathcal{C}_j|$ for all j's in the same equivalence class must also be smaller than ρn. So, we can partition the set $\{1,\dots,\kappa\}$ into sets A and B in such a way that equivalent indices belong to the same set and $|\sum_{a\in A}|\mathcal{C}_a| - \sum_{b\in B}|\mathcal{C}_b|| \le 2\rho n$. Since $\sum_j |\mathcal{C}_j| \ge 100\rho n$, we obtain that $\sum_{a\in A}|\mathcal{C}_a|$ and $\sum_{b\in B}|\mathcal{C}_b| \ge 10\rho n$, and, by construction of the equivalence relation $\sim$, $\bigcup_{a\in A}\mathcal{C}_a$ is not connected to $\bigcup_{b\in B}\mathcal{C}_b$ through $\operatorname{supp}\xi^{\mathrm{sp}}$. This shows (6.106).

For subsets F and F' of V, we use the notation $F \overset{\mathrm{sp}}{\leftrightarrow} F'$ to denote the event

$$\big\{F \text{ and } F' \text{ are not connected in } \operatorname{supp}\xi^{\mathrm{sp}}\big\}. \tag{6.107}$$

By (6.106) and $\mathcal{K} \subseteq \{\kappa < \infty\}$,

$$\mathsf{P}\otimes\mathcal{Q}\big[\big|\mathcal{C}_{\max}(\xi^{\mathrm{sp}})\big| < \rho n\big] \le \mathsf{P}\big[\mathcal{S}^c\big] + \mathsf{P}\big[\mathcal{K}^c\big] + \mathsf{E}\big[\mathbf{1}_{\mathcal{K}\cap\mathcal{S}}\mathcal{Q}[\mathcal{P}]\big], \tag{6.108}$$

where we have used that $\mathcal{S},\mathcal{K}\subset\Omega$ and Fubini's theorem. For the sake of clarity, let us recall that $\mathcal{Q}$ is a measure on $\{0,1\}^{M_{u_n}}$ and emphasize that the $\mathcal{Q}$-probability in this last expression is computed for Y^i and $Z^{i,j}$ fixed. On $\mathcal{K}$, there are at most $2^{\lceil n^{1-v_{u'}(1+\varepsilon)^\beta}\rceil}$ ways to partition $\{1,\dots,\kappa\}$ into A, B. Hence, on $\mathcal{K}$, using the union bound,

$$\mathcal{Q}[\mathcal{P}] \le 2^{\lceil n^{1-v_{u'}(1+\varepsilon)^\beta}\rceil}\sup \mathcal{Q}[\mathcal{A} \overset{\mathrm{sp}}{\leftrightarrow} \mathcal{B}], \tag{6.109}$$

where the supremum is taken over all partitions of $\{1,\dots,\kappa\}$ as in (6.105) and $\mathcal{A}$, $\mathcal{B}$ are defined in (6.105), too. By increasing the range of the supremum we deduce that the following estimate holds uniformly on the event $\mathcal{K}$,

$$\mathcal{Q}[\mathcal{P}] \le 2^{\lceil n^{1-v_{u'}(1+\varepsilon)^\beta}\rceil}\sup_{F,F'\subseteq V:|F|,|F'|\ge 10\rho n}\mathcal{Q}\big[F \overset{\mathrm{sp}}{\leftrightarrow} F'\big]. \tag{6.110}$$

We will now find a bound on the event on the right-hand side, valid uniformly on the event $\mathcal{S}$. To this end, fix Y^i and $Z^{i,j}$ such that $\mathcal{S}$ holds, as well as subsets F and F' of V containing at least $10\rho n$ vertices. Using the expansion property of the graph G, see (2.11), and the Max-flow Min-cut Theorem, we can find a collection of at least $c_5 n$ disjoint paths in V joining the sets F and F' for some constant $c_5>0$. We call these paths *connections*. Since, on $\mathcal{S}$, $\sum_{i<M_U} N^i$ is smaller or equal to $2u_\star n$ (see (4.10) and (6.103)), we can extract from this collection a sub-collection $\mathbf{C}$ such that $|\mathbf{C}| = \frac{1}{2}c_5 n$, and such that all connections in $\mathbf{C}$ intersect at most $\lfloor 4u_\star/c_5\rfloor$ segments Y^i.

We next want to prove that with high probability, at least one of these connections only intersects Y^i's that do not appear in ξ^{sp}, using again the concentration inequality from [16], Lemma 1.2. To this end, we define the function g by

$$g(R_1,\dots,R_{M_{u_n}}) = \big|\big\{\zeta\in\mathbf{C}\colon \text{ for all } i, \text{ either } \zeta\cap Y^i = \varnothing \text{ or } R_i = 1\big\}\big|. \tag{6.111}$$

The probability that all of the at most $\lfloor 4u_\star/c_5\rfloor$ Y^i's intersecting a given $\zeta\in\mathbf{C}$ have an index i with $R_i=1$ is at least $q^{\lfloor 4u_\star/c_5\rfloor}$. So, for some $c_6 = c_6(\alpha_1,\alpha_2,u)$, $c_7>0$,

$$E^{\mathcal{Q}}[g] \ge \frac{1}{2}c_5 q^{\lfloor 4u_\star/c_5\rfloor} n =: 2c_6 n^{1-c_7\delta}. \tag{6.112}$$

Changing one segment Y^i can change the value of g by at most $2n^\gamma$ on $\mathcal{S}$. Therefore, by Lemma 1.2 in [16],

$$\mathcal{Q}\big[g < c_6 n^{1-c_7\delta}\big] \le 2\exp\left\{\frac{-c_6^2 n^{2(1-c_7\delta)}}{n^{1-\gamma}n^{2\gamma}}\right\}. \tag{6.113}$$

If $g \ge c_6 n^{1-c_7\delta}$, then there are at least $c_6 n^{1-c_7\delta}$ disjoint connections in $\mathbf{C}$ linking F and F' and only using vertices in $\operatorname{supp}\xi^{\mathrm{sp}}$. In accordance with (6.87), we now choose δ such that $2c_7\delta<\gamma$. Then, since on $\mathcal{S}$ the total length of the bridges $\sum_{i,j}N^{i,j} \le \ln^5 n\cdot n^{1-\gamma}$, the events $\{g\ge c_6 n^{1-c_7\delta}\}$ and $\mathcal{S}$ imply that at least one connection in $\mathbf{C}$ is contained in $\operatorname{supp}\xi^{\mathrm{sp}}$, for $n\ge c_{u,\delta}$. Hence, (6.113) implies that uniformly on the event $\mathcal{S}$,

$$\sup_{F,F'\subset V:|F|,|F'|\ge 10\rho n}\mathcal{Q}\big[F \overset{\mathrm{sp}}{\leftrightarrow} F'\big] \le 2\exp\big(-c_6^2 n^{1-\gamma-2c_7\delta}\big). \tag{6.114}$$

Inserting this estimate into (6.110), noting that

$$1-\gamma-2c_7\delta > 1-2\gamma \overset{(6.11)}{=} 1-v_{u'(1+\varepsilon)}\beta, \tag{6.115}$$

we find that, uniformly on $\mathcal{K}\cap\mathcal{S}$,

$$\mathcal{Q}[\mathcal{P}] \le \exp\bigl(-c_{u,\delta}n^{1-2\gamma}\bigr). \tag{6.116}$$

Using this estimate, together with the bound (6.102) on $\mathsf{P}[\mathcal{K}^c]$ and the bound (6.104) on $\mathsf{P}[\mathcal{S}^c]$, in (6.108), we find (6.99) for $n\ge 1$ by possibly adjusting the constants. This concludes the proof of Theorem 1.2. □

7. Uniqueness of the giant component

This section contains the proof of Theorem 1.3, that is of the uniqueness of the giant component. More precisely, we show that for any choice of $\kappa>0$ and $u<u_\star$, with a high probability, the second largest component of the vacant set $\mathcal{V}_n^u$ is smaller than κn. The sprinkling is again the major ingredient of the proof. This time, however, we will really use the fact that $u_n-u<n^{-\delta}$.

Heuristically, our argument runs as follows. We will show that any component of $\mathcal{V}_n^u$ of size at least κn should contain at least $\kappa n/2$ vertices that were included in clusters of size at least $n^{v_u\beta/2}$ of the vacant set left by segments at level u_n. Hence, in order to have $|\mathcal{C}_{\mathrm{sec}}|\ge\kappa n$, there should be two groups of such vertices which do not get connected after the sprinkling. A small extension of the proof of the last section then shows that this happens with a small probability.

Proof of Theorem 1.3. We choose γ,β as in (6.11) and recall from (6.87) the notation $u_n=(u+n^{-\delta})\wedge((u+u_\star)/2)$, where $2\delta<1-\gamma$. By decreasing δ, we can also assume that $\delta<\gamma$. Recall also that the vacant set left by segments ξ_{u_n} was defined in (6.14) (with u_n replaced by u).

We divide the vertices of G into three sets. The vertex $x\in V$ is called *small*, if

$$\mathtt{tx}\bigl(B(x,5\beta\,\mathrm{ld}\,n)\bigr)=0,\qquad \bigl|\mathcal{C}_x(\xi_{u_n})\bigr|\le \mathrm{ld}^2 n \quad\text{and}\quad \mathcal{C}_x(\xi_{u_n})\subset B(x,\beta\,\mathrm{ld}\,n). \tag{7.1}$$

It is called *proper* (cf. Definition 6.7), if

$$\mathtt{tx}\bigl(B(x,5\beta\,\mathrm{ld}\,n)\bigr)=0 \quad\text{and}\quad \text{(i), (ii) of (6.56) hold with } h=c_3 \text{ of Lemma 6.8.} \tag{7.2}$$

It is called *bad* otherwise. We will use $\mathcal{B}$ to denote the set of bad vertices.

The next lemma shows that the set $\mathcal{B}$ is small. The lemma should be viewed as an analogue to a non-existence of intermediate components in the Bernoulli percolation case.

Lemma 7.1. *There exists a function $g(n)$ such that $\lim_{n\to\infty} g(n)/n=0$ and*

$$\lim_{n\to\infty}\mathsf{P}\bigl[|\mathcal{B}|\ge g(n)\bigr]=0. \tag{7.3}$$

We postpone the proof of the Lemma 7.1 and proceed with the proof of Theorem 1.3. First, we add long-range bridges to the configuration ξ_{u_n}, that is we define ξ'_{u_n} as in (6.53). We must be careful to see that these bridges do not destroy the components of proper vertices in $\operatorname{supp}\xi_{u'}$. To this end we collect a family Γ of strings (that is of self-avoiding paths of length $\lfloor\beta\,\mathrm{ld}\,n\rfloor$) as in the proof of Lemma 6.9. This collection contains $\lfloor c_3 n^{\beta v_+}\rfloor$ distinct strings starting at x for any proper vertex x, as before. In particular, this implies that $|\Gamma|\le c_3 n^{1+\beta v_+}$. We can show, as below (6.75), that the number S_y of strings intersecting a given $y\in V$ satisfies $S_y\le c_u(\log^{c_u} n)n^{(5/4)v_-\beta}$. Let now $\mathcal{S}$ be the event that all segments and bridges are not too long, as in (6.103). On $\mathcal{S}$, the number of strings that are intersected by bridges is thus at most $c_u(\log^{c'_u} n)n^{1-\gamma+(5/4)v_-\beta}$. We declare a proper vertex *bad proper*, it at least half of the strings starting at this vertex is intersected by the bridges. Obviously, on $\mathcal{S}$, the set $\mathcal{BP}$ of bad proper vertices must satisfy

$|\mathcal{BP}|\frac{c_3}{2}n^{\beta v_+} \le c_u(\log^{c'_u} n)n^{1-\gamma+(5/4)v_-\beta}$. With (6.10) and (6.11), it follows that

$$|\mathcal{BP}| = \mathrm{o}(n) \quad \text{as } n \to \infty, \text{ on } \mathcal{S}. \tag{7.4}$$

On the other hand, the remaining proper vertices, that we call *large*, are starting vertices of at least $\frac{1}{2}n^{v_+\beta}$ strings which are not intersected by the bridges. Hence, if x is large, it is contained in a component of ξ'_{u_n} of size at least $\frac{1}{2}n^{v_+\beta}$. This implies that

$$\text{on } \mathcal{S}, \text{ the number of components of } \operatorname{supp}\xi'_{u_n} \text{ that contain a large vertex is at most } cn^{1-\beta v_+}. \tag{7.5}$$

We now perform the sprinkling as in Section 6. Recall that on the probability space $(\{0,1\}^{M_{u_n}}, \mathcal{Q})$ we have defined i.i.d. random variables $(R_k)_{0\le k<M_{u_n}}$ with success probability $q = n^{-2\delta}$ (cf. (6.88)), the number of remaining segments I (cf. (6.89)) and the good event $\mathcal{G}$ (cf. (6.90)). We have then constructed the sprinkled configuration ξ^{sp} (cf. (6.92)). Using Lemma 4.1, then the estimate (6.95), as $n \to \infty$,

$$\begin{aligned} P\big[\big|\mathcal{C}^u_{\mathrm{sec}}\big| \ge \kappa n\big] &\le \mathcal{Q}^{un}\big[\big|\mathcal{C}^u_{\mathrm{sec}}\big| \ge \kappa n\big] + \mathrm{o}(1) \\ &\le \mathsf{P}\otimes\mathcal{Q}\big[\big|\mathcal{C}_{\mathrm{sec}}\big(\xi^{\mathrm{sp}}\big)\big| \ge \kappa n, \mathcal{G}, \mathcal{S}, |\mathcal{B}| \le g(n)\big] \\ &\quad + \mathcal{P}\big[\mathcal{G}^c\big] + \mathsf{P}\big[\mathcal{S}^c\big] + \mathsf{P}\big[|\mathcal{B}| > g(n)\big] + \mathrm{o}(1) \\ &\overset{(6.94),(6.104),(7.3)}{\le} \mathsf{P}\otimes\mathcal{Q}\big[\big|\mathcal{C}_{\mathrm{sec}}\big(\xi^{\mathrm{sp}}\big)\big| \ge \kappa n, \mathcal{G}, \mathcal{S}, \big\{|\mathcal{B}| \le g(n)\big\}\big] + \mathrm{o}(1). \end{aligned} \tag{7.6}$$

In order to estimate the term on the right-hand side, we claim that for any $\kappa > 0$,

$$\begin{aligned} &\text{on } \mathcal{S}\cap\mathcal{G}\cap\{|\mathcal{B}| \le g(n)\}, \text{ any component of } \operatorname{supp}\xi^{\mathrm{sp}} \text{ of size } \ge \kappa n \\ &\text{contains at least } \kappa n/2 \text{ large vertices, for } n \ge c_{\kappa,u,\delta}. \end{aligned} \tag{7.7}$$

Indeed, recall that we have divided the vertices in $\operatorname{supp}\xi'_{u_n}$ into small, bad, bad proper and large vertices. The vertices in $\operatorname{supp}\xi^{\mathrm{sp}}$ consist of these four sets, and the set $\{x \in V\colon \xi'_{u_n}(x) = 0, \xi^{\mathrm{sp}}(x) = 1\}$. Let us call all the vertices in this last set *sprinkled* vertices. Suppose now that the event $\mathcal{S}\cap\mathcal{G}\cap\{|\mathcal{B}| \le g(n)\}$ occurs. Then by definition of $\mathcal{S}$ and $\mathcal{G}$, the number of sprinkled vertices is at most

$$(M_{u_n} - I)2n^{\gamma} + cn^{1-\gamma}\ln^5 n \le (M_{u_n} - M_u)2n^{\gamma} + cn^{1-\gamma}\ln^5 n \le c_{u,\delta}n^{1-\delta}, \tag{7.8}$$

because $M_{u_n} - M_u \le c_{u,\delta}n^{1-\gamma-\delta}$ and $\delta < \gamma$. Consider now any component A of size κn of $\operatorname{supp}\xi^{\mathrm{sp}}$. Then the number of vertices in A that are either bad, bad proper or sprinkled is at most $|\mathcal{B}| + |\mathcal{BP}| + c_{u,\delta}n^{1-\delta} = \mathrm{o}(n)$, by definition of $g(n)$ and (7.4), the remaining vertices being either small or large. By definition of small vertex, all small vertices in $\operatorname{supp}\xi'_{u_n}$ belong to components of size at most $\mathrm{ld}^2 n$. Any of the at most $c_{u,\delta}n^{1-\delta}$ sprinkled vertices can merge at most $(d-1)$ such components, so the maximum number of small vertices belonging to the same component of $\operatorname{supp}\xi^{\mathrm{sp}}$ is bounded by $c_{u,\delta}(\mathrm{ld}^2 n)n^{1-\delta}$, which is less than $\mathrm{o}(n)$, too. Hence, for $n \ge c_{\kappa,u,\delta}$, at least $\kappa n/2$ of the vertices in A are large, proving (7.7).

Let $\mathcal{L}$ be the set of large vertices. Defining $\mathcal{P}'$ to be the event

$$\mathcal{P}' = \big\{\text{there are } A, B \subset \mathcal{L} \text{ such that } |A| \ge \kappa n/2, |B| \ge \kappa n/2 \text{ and } A \overset{\mathrm{sp}}{\leftrightarrow} B\big\}, \tag{7.9}$$

we obtain from (7.7), for $n \ge c_{\kappa,u,\delta}$,

$$\mathsf{P}\otimes\mathcal{Q}\big[\big|\mathcal{C}_{\mathrm{sec}}\big(\xi^{\mathrm{sp}}\big)\big| \ge \kappa n, \mathcal{G}, \mathcal{S}, \big\{|\mathcal{B}| \le g(n)\big\}\big] \le \mathsf{P}\otimes\mathcal{Q}\big[\mathcal{P}', \mathcal{S}\big] = \mathsf{E}\big[1_{\mathcal{S}}\mathcal{Q}\big[\mathcal{P}'\big]\big], \tag{7.10}$$

which is essentially equivalent to the right-hand side of (6.108). Note also that since $\operatorname{supp}\xi'_{u_n} \subseteq \operatorname{supp}\xi^{\mathrm{sp}}$, $\mathcal{P}'$ equals

$$\Big\{\text{there are } A, B \subset \mathcal{L} \text{ such that } |A|, |B| \ge \kappa n/2 \text{ and } \bigcup_{a\in A}\mathcal{C}_a\big(\xi'_{u_n}\big) \overset{\mathrm{sp}}{\leftrightarrow} \bigcup_{b\in B}\mathcal{C}_b\big(\xi'_{u_n}\big)\Big\}.$$

We should now bound the number of possible choices for the unions in the equation above. By (7.5), there are at most $n^{1-\beta v_+}$ sets of the form $\mathcal{C}_x(\xi'_{u_n})$ with $x \in \mathcal{L}$, so there are at most $2^{2n^{1-\beta v_+}}$ choices for the unions in the equation above. Hence,

$$\mathsf{E}\big[1_{\mathcal{S}}\mathcal{Q}\big[\mathcal{P}'\big]\big] \le 2^{2n^{1-\beta v_+}} \sup_{\mathcal{S}} \sup_{F,F'\subset V:|F|,|F'|\ge \kappa n/2} \mathcal{Q}\big[F \overset{\text{sp}}{\leftrightarrow} F'\big]. \tag{7.11}$$

Repeating the argument from (6.110) to (6.114) and choosing δ small enough, we infer that the right-hand side of (7.11) tends to zero as n tends to infinity. With (7.6) and (7.10), this completes the proof of Theorem 1.3. □

We now prove the lemma we used in previous proof. Due to Proposition 6.3, this proof will be reduced to estimates on a branching process.

Proof of Lemma 7.1. Let $x \in V$ be an arbitrary vertex and let $B_x = B(x, \beta \operatorname{ld} n)$, $B'_x = B(x, 5\beta \operatorname{ld} n)$. We write $v_+ = v_{u(1+\varepsilon)}$, $v_- = v_{u(1-\varepsilon)}$, see (6.8) for the notation. It is easy to see that $\mathcal{B} \subset \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_4$, where

$$\begin{aligned}
\mathcal{B}_1 &= \big\{x \in V\colon \mathtt{tx}\big(B'_x\big) > 0\big\},\\
\mathcal{B}_2 &= \big\{x \in V\colon \mathtt{tx}\big(B'_x\big) = 0, \big|\mathcal{C}_x^{\beta \operatorname{ld} n}(\xi_{u_n})\big| < c_3 n^{v_+\beta} \text{ and } \mathcal{C}_x(\xi_{u_n}) \not\subset B_x\big\},\\
\mathcal{B}_3 &= \big\{x \in V\colon \mathtt{tx}\big(B'_x\big) = 0, \big|\mathcal{C}_x^{\beta \operatorname{ld} n}(\xi_{u_n})\big| < c_3 n^{v_+\beta} \text{ and } \big|\mathcal{C}_x(\xi_{u_n})\big| > \operatorname{ld}^2 n\big\},\\
\mathcal{B}_4 &= \big\{x \in V\colon \mathtt{tx}\big(B'_x\big) = 0, \text{(ii) of (6.56) does not hold with } h = c_3\big\}.
\end{aligned} \tag{7.12}$$

By Lemma 6.1 we have $|\mathcal{B}_1| \le g(n)$ for a sequence g which decays as in the statement, deterministically. Further, by (6.63), we have $\mathsf{P}[x \in \mathcal{B}_4] \le cn^{-\beta}$. Therefore, using the Markov inequality, there is a sequence $g(n)$ such that $g(n)/n$ tends to zero, such that $\lim_{n\to\infty} \mathsf{P}[|\mathcal{B}_4| \ge g(n)] = 0$.

It remains to control $\mathcal{B}_2$, $\mathcal{B}_3$. We choose any $\varepsilon > 0$ small enough such that (6.9) is satisfied, noting that these two constraints allow us to make $\varepsilon > 0$ even smaller. For $n \ge c_{\delta,u}$, we then have $u_n \in (u, u(1+\varepsilon/4))$. Hence, the random sets $\mathcal{C}^{u(1\pm\varepsilon/2)}$ constructed in Proposition 6.3 (with ε replaced by $\varepsilon/2$) dominate the component $\mathcal{C}_y(1_B \cdot \xi_{u_n})$ from above and from below with probability at least $1 - c_{u,\delta,\varepsilon} n^{-2\beta}$, for $n \ge c_{u,\delta,\varepsilon}$. Recall also that $\mathcal{C}^{u(1\pm\varepsilon/2)}$ are distributed as $\mathcal{C}_o$ under $\mathbb{Q}^\star_{u(1\pm\varepsilon/2)}$. Let Z_k^- be the branching process description of $\mathcal{C}^{u(1+\varepsilon/2)}$, that is $Z_k^- = |\{y \in \mathcal{C}^{u(1+\varepsilon/2)}, \operatorname{dist}(x, y) = k\}|$. Similarly, let Z_k^+ be such a description of $\mathcal{C}^{u(1-\varepsilon/2)}$. By our choice of parameters and Lemma 6.2, both Z^+ and Z^- are supercritical branching processes. We use T_+, T_- to denote their extinction times, ϕ_+, ϕ_- their offspring generating function, and q_+, q_- their extinction probabilities. Observe that

$$\lim_{\varepsilon\to 0} q_- - q_+ = 0. \tag{7.13}$$

Set $r = \lfloor \beta \operatorname{ld} n \rfloor$, $N_n = c_3 n^{\beta v_+}$. Using Proposition 6.3, we get

$$\begin{aligned}
\mathsf{P}[x \in \mathcal{B}_2] - c_{u,\delta,\varepsilon} n^{-2\beta} &\le \mathsf{P}\big[Z_r^- \le N_n, Z_r^+ \ge 1\big]\\
&\le \mathsf{P}\big[1 \le Z_r^- \le N_n\big] + \mathsf{P}\big[0 = Z_r^- < Z_r^+\big].
\end{aligned} \tag{7.14}$$

Since $v_\cdot$ is strictly decreasing, $N_n/n^{\beta v_{u(1+\varepsilon/2)}} \to 0$. Hence, the first term on the right-hand side is the probability that the branching process is not extinct at generation r, but is much smaller than its typical size $n^{\beta v_{u(1+\varepsilon/2)}}$. This probability tends to 0 as $n \to \infty$, using e.g. Theorems 6.1, 6.2 in Chapter I, p. 9, of [5]. Using the fact that the generating function of $Z_r^\pm$ is the rth iteration $\phi_\pm^{(r)}$ of ϕ, we get that

$$\limsup_{n\to\infty} \mathsf{P}\big[0 = Z_r^- < Z_r^+\big] = \limsup_{r\to\infty} \phi_-^{(r)}(0) - \phi_+^{(r)}(0) = q_- - q_+. \tag{7.15}$$

Here we have used that $q_\pm$ is the attractive fixed-point of $\phi_\pm$. Using (7.13), this can be made arbitrarily small by choosing ε small. Inserting this back into (7.14), we get $\lim_{n\to\infty} \mathsf{P}[x \in \mathcal{B}_2] = 0$. This implies, using the Markov inequality again, that $\lim_{n\to\infty} \mathsf{P}[|\mathcal{B}_2| \ge g(n)] = 0$ for some g as in the statement.

Similarly we have,

$$\begin{aligned}\mathsf{P}[x \in \mathcal{B}_3] - cn^{-2\beta} &\le \mathsf{P}\left[Z_r^- \le N_n, \sum_{k=0}^{\infty} Z_k^+ \ge \mathrm{ld}^2 n\right] \\ &\le \mathsf{P}\big[Z_r^- \le N_n, T_- = \infty\big] + \mathsf{P}\left[\sum_{k=0}^{\infty} Z_k^+ \ge \mathrm{ld}^2 n, T_- < \infty\right] \\ &\le \mathsf{P}\big[1 \le Z_r^- \le N_n\big] + \mathsf{P}\left[\sum_{k=0}^{\infty} Z_k^+ \ge \mathrm{ld}^2 n, T_+ < \infty\right] + \mathsf{P}[T_- < T_+ = \infty]. \qquad (7.16)\end{aligned}$$

The first probability on the right-hand side tends to 0, as in the previous argument. By [5], Theorem 12.3 in Chapter I, p. 52, conditioned on $T_+ < \infty$, Z^+ has the law of a sub-critical branching process. Using this claim it is easy to show that the second probability in (7.16) tends to zero. The third probability can be made arbitrarily small by choosing ε small, by using (7.13) again. This then implies that $\lim_{n\to\infty} \mathsf{P}[|\mathcal{B}_3| \ge g(n)] = 0$ for an appropriately chosen g, as in the previous case. This completes the proof of the lemma. □

Acknowledgements

The authors wish to thank Itai Benjamini for proposing the study of the vacant set on expanders and are indebted to Alain-Sol Sznitman for helpful discussions. This work was started when the third author was at ETH Zurich.

References

[1] D. J. Aldous and M. Brown. Inequalities for rare events in time-reversible Markov chains. II. *Stochastic Process. Appl.* **44** (1993) 15–25. MR1198660
[2] D. J. Aldous and J. A. Fill. Reversible Markov chains and random walks on graphs. Available at http://www.stat.berkeley.edu/~aldous/RWG/book.html.
[3] N. Alon, I. Benjamini and A. Stacey. Percolation on finite graphs and isoperimetric inequalities. *Ann. Probab.* **32** (2004) 1727–1745. MR2073175
[4] K. B. Athreya. Large deviation rates for branching processes. I. Single type case. *Ann. Appl. Probab.* **4** (1994) 779–790. MR1284985
[5] K. B. Athreya and P. E. Ney. *Branching Processes*. *Die Grundlehren der Mathematischen Wissenschaften* **196**. Springer, New York, 1972. MR0373040
[6] I. Benjamini and A.-S. Sznitman. Giant component and vacant set for random walk on a discrete torus. *J. Eur. Math. Soc. (JEMS)* **10** (2008) 133–172. MR2349899
[7] C. Borgs, J. T. Chayes, R. van der Hofstad, G. Slade and J. Spencer. Random subgraphs of finite graphs. I. The scaling window under the triangle condition. *Random Structures Algorithms* **27** (2005) 137–184. MR2155704
[8] A. Broder and E. Shamir. On the second eigenvalue of random regular graphs. In *28th Annual Symposium on Foundations of Computer Science* 286–294. IEEE Comput. Soc. Press, Washington, DC, 1987.
[9] A. Dembo and A.-S. Sznitman. On the disconnection of a discrete cylinder by a random walk. *Probab. Theory Related Fields* **136** (2006) 321–340. MR2240791
[10] R. Durrett. *Probability: Theory and Examples*, 2nd edition. Duxbury Press, Belmont, CA, 1996.
[11] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960) 17–61. MR0125031
[12] J. Friedman. On the second eigenvalue and random walks in random d-regular graphs. *Combinatorica* **11** (1991) 331–362. MR1137767
[13] J. Friedman. A proof of Alon's second eigenvalue conjecture and related problems. *Mem. Amer. Math. Soc.* **195** (2008) viii + 100. MR2437174
[14] E. Lubetzky and A. Sly. Cutoff phenomena for random walks on random regular graphs. *Duke Math. J.* **153** (2010) 475–510. Available at arXiv:0812.0060. MR2667423
[15] A. Lubotzky, R. Phillips and P. Sarnak. Ramanujan graphs. *Combinatorica* **8** (1988) 261–277. MR0963118
[16] C. McDiarmid. On the method of bounded differences. In *Surveys in Combinatorics, 1989 (Norwich, 1989)* 148–188. *London Math. Soc. Lecture Note Ser.* **141**. Cambridge Univ. Press, Cambridge, 1989. MR1036755
[17] A. Nachmias and Y. Peres. Critical percolation on random regular graphs. *Random Structures Algorithms* **36** (2010) 111–148. Available at arXiv:0707.2839. MR2583058
[18] B. Pittel. Edge percolation on a random regular graph of low degree. *Ann. Probab.* **36** (2008) 1359–1389. MR2435852
[19] L. Saloff-Coste. Lectures on finite Markov chains. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1996)* 301–413. *Lecture Notes in Math.* **1665**. Springer, Berlin, 1997. MR1490046

[20] V. Sidoravicius and A.-S. Sznitman. Percolation for the vacant set of random interlacements. *Comm. Pure Appl. Math.* **62** (2009) 831–858. MR2512613

[21] A.-S. Sznitman. A lower bound on the critical parameter of interlacement percolation in high dimension. *Probab. Theory Related Fields*. To appear (2011).

[22] A.-S. Sznitman. On the domination of random walk on a discrete cylinder by random interlacements. *Electron. J. Probab.* **14** (2009) 1670–1704. MR2525107

[23] A.-S. Sznitman. Random walks on discrete cylinders and random interlacements. *Probab. Theory Related Fields* **145** (2009) 143–174. MR2520124

[24] A.-S. Sznitman. Upper bound on the disconnection time of discrete cylinders and random interlacements. *Ann. Probab.* **37** (2009) 1715–1746. MR2561432

[25] A.-S. Sznitman. Vacant set of random interlacements and percolation. *Ann. of Math. (2)* **171** (2010) 2039–2087. MR2680403

[26] A. Teixeira. Interlacement percolation on transient weighted graphs. *Electron. J. Probab.* **14** (2009) 1604–1628. MR2525105

[27] A. Teixeira and D. Windisch. On the fragmentation of a torus by random walk. *Commun. Pure Appl. Math.* To appear (2011). Available at arXiv:1007.0902.

[28] D. Windisch. Random walk on a discrete torus and random interlacements. *Electron. Commun. Probab.* **13** (2008) 140–150. MR2386070