CRITICAL WINDOW FOR THE VACANT SET LEFT BY RANDOM WALK ON RANDOM REGULAR GRAPHS

JIŘÍ ČERNÝ[1] AND AUGUSTO TEIXEIRA[2]

Abstract. We consider the simple random walk on a random d-regular graph with n vertices, and investigate percolative properties of the set of vertices not visited by the walk until time un, where $u > 0$ is a fixed positive parameter. It was shown in [ČTW11] that this so-called vacant set exhibits a phase transition at $u = u_\star$: there is a giant component if $u < u_\star$ and only small components when $u > u_\star$. In this paper we show the existence of a critical window of size $n^{-1/3}$ around $u_\star$. In this window the size of the largest cluster is of order $n^{2/3}$.

1. Introduction

The study of percolative properties of the vacant set left by a random walk on finite graphs was initiated by Benjamini and Sznitman [BS08] for the case of random walk on a high-dimensional discrete torus $(\mathbb{Z}/N\mathbb{Z})^d$. In [BS08] it is proved that if the random walk runs up to time uN^d, where u is a small constant, the vacant set has a giant component with volume of order N^d, asymptotically as N grows. On the other hand, if instead we consider a large constant u, all components of the vacant set have a volume of order at most $\log^\lambda N$ (for some $\lambda > 0$) as was proved in [TW10]. This shows the existence of two distinct phases for the connectivity of the vacant set on the torus as u varies. However, the above mentioned works leave several open questions, such as whether the transition between these two phases happens abruptly at a given critical threshold and, if this is the case, how does the vacant set behave at the critical point?

Such problems are much better understood when instead of the torus one considers random d-regular graphs or d-regular large-girth expanders on n vertices. In this case, when random walks runs up to time un, it is known that the vacant set exhibits a sharp phase transition [ČTW11]: the size of the largest connected component of the vacant set drops abruptly from order n to order $\log n$ at a computable critical value $u_\star$. In this paper we explore more closely this phase transition, in particular we prove that the size of the largest component of the vacant set exhibits a double-jump, similar to that observed in Erdős-Rényi random graphs.

Let us now give a precise definition of the model. Let $\mathcal{G}_{n,d}$ be the set of all non-oriented d-regular simple graphs with n vertices (here and later we tacitly assume that nd is even). Let $\mathbb{P}_{n,d}$ be the uniform probability distribution on $\mathcal{G}_{n,d}$. For any graph $G = (V, \mathcal{E})$ let P^G denote the canonical law on the Skorokhod space $D([0,\infty), V)$ of a continuous-time simple random walk on G started from the uniform distribution. We use $(X_t)_{t\ge 0}$ to denote the

Date: 10 January 2011.

[1] Department of Mathematics, ETH Zurich, Raemistrasse 101, 8092 Zurich, Switzerland.

[2] Department of Mathematics and their Applications, Ecole Normale Supérieure, 45 rue d'Ulm, F-75230 Paris, France.

This research was partially supported by the AXA Research Fund Fellowship. A.T. thanks the Forschungsinstitut für Mathematik (FIM) of ETH Zurich, where a part of this work has been completed, for its hospitality.

canonical coordinate process. For a fixed parameter $u \geq 0$, we define the vacant set as the set of all vertices not visited by the random walk before the time $u|V|$,

$$\mathcal{V}^u = \mathcal{V}^u_G = V \setminus \{X_t : 0 \leq t \leq u|V|\}. \tag{1.1}$$

We use $\mathcal{C}^u_{\max}$ to denote the maximal connected component of the subgraph of G induced by $\mathcal{V}^u$. The vacant set and in particular its maximal connected component are the main objects of investigation in this paper.

As proved in [ČTW11], the phase transition in the connectivity of the vacant set occurs at the value $u_\star$ given by

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

and can be described as follows: Let G_n be a graph distributed according to $\mathbb{P}_{n,d}$. Then with $\mathbb{P}_{n,d}$-probability tending to one as $n \to \infty$:

Super-critical phase: For any $u < u_\star$ and $\sigma > 0$ there exist ρ and c depending on u, σ, and d, such that

$$P^{G_n}[|\mathcal{C}^u_{\max}| \geq \rho n] \geq 1 - cn^{-\sigma}. \tag{1.3}$$

Sub-critical phase: For any $u > u_\star$ and $\sigma > 0$ there exist K and c depending on u, σ, and d, such that

$$P^{G_n}[|\mathcal{C}^u_{\max}| \leq K \log n] \geq 1 - cn^{-\sigma}. \tag{1.4}$$

In this paper we study the behaviour of the vacant set in the vicinity of the critical point. The main results of this paper are the following two theorems. In their statement we use $\mathbf{P}_{n,d}$ to denote the *annealed* measure

$$\mathbf{P}_{n,d}(\cdot) = \int P^G(\cdot)\mathbb{P}_{n,d}(\mathrm{d}G). \tag{1.5}$$

We say that an event A occurs $\mathbf{P}_{n,d}$-asymptotically almost surely (or simply $\mathbf{P}_{n,d}$-a.a.s.) if $\lim_{n\to\infty} \mathbf{P}_{n,d}(A) = 1$.

Theorem 1.1 (Critical window)**.** *Let $(u_n)_{n\geq 1}$ be a sequence satisfying*

$$|n^{1/3}(u_n - u_\star)| \leq \lambda < \infty \quad \text{for all } n \text{ large enough.} \tag{1.6}$$

Then for every $\varepsilon > 0$ there exists $A = A(\varepsilon, d, \lambda)$ such that for all n large enough

$$\mathbf{P}_{n,d}[A^{-1}n^{2/3} \leq |\mathcal{C}^{u_n}_{\max}| \leq An^{2/3}] \geq 1 - \varepsilon. \tag{1.7}$$

If u_n is not in the critical window, then the maximal connected component behaves differently:

Theorem 1.2. *(a) When $(u_n)_{n\geq 1}$ satisfies*

$$u_\star - u_n \xrightarrow{n\to\infty} 0, \qquad \text{and} \qquad \omega_n := n^{1/3}(u_\star - u_n) \xrightarrow{n\to\infty} \infty, \tag{1.8}$$

then for $v_n = 2n^{2/3}\omega_n \frac{d-2}{(d-1)^2} e^{-u_\star(d-2)/(d-1)}$ and for every $\varepsilon > 0$

$$\left| \frac{|\mathcal{C}^{u_n}_{\max}|}{v_n} - 1 \right| \leq \varepsilon \qquad \mathbf{P}_{n,d}\text{-a.a.s.} \tag{1.9}$$

(b) When $(u_n)_{n\geq 1}$ satisfies

$$u_\star - u_n \xrightarrow{n\to\infty} 0, \qquad \text{and} \qquad \omega_n := n^{1/3}(u_\star - u_n) \xrightarrow{n\to\infty} -\infty, \tag{1.10}$$

then for every $\varepsilon > 0$ there exists $B = B(\varepsilon) > 0$, such that for all n large enough

$$\mathbf{P}_{n,d}\big[|\mathcal{C}^{u_n}_{\max}| \le Bn^{2/3}|\omega_n|^{-1/2}\big] \ge 1-\varepsilon. \tag{1.11}$$

The above theorems confirm that the maximal connected component of the vacant set behaves similarly as the largest connected cluster of the Bernoulli percolation on random regular graphs, see [ABS04, NP10, Pit08]. Remark that the upper bound in part (b) of Theorem 1.2 seems to be non-optimal. The result is however sufficient to confirm that the width of the window is $n^{-1/3}$. To improve such a statement, it is necessary to obtain better estimate in Theorem 3.2(ii) that we quote below.

The methods of this paper are largely inspired by the recent article by Cooper and Frieze [CF10], where the authors develop a new technique to prove (1.3) and (1.4). This technique is very specific to deal with random regular graphs, in contrast with the results of [ČTW11] which hold for a more general class of graphs, including e.g. large girth expanders. In the present article we extend the methods in [CF10] to the critical case. We believe that obtaining such an extension from the techniques in [ČTW11] should be rather difficult.

The crucial observation of [CF10], allowing for a very elegant proof of (1.3), (1.4) for random regular graphs is the following. Under the annealed measure (1.5), given the information about the graph discovered by the random walk up to time $u|V|$, the subgraph of G induced by the vacant set $\mathcal{V}^u$ is distributed as a random graph uniformly chosen within the set of graphs with a given (random) degree sequence, see Proposition 3.1 below.

The paper [CF10] further uses the fact that the behaviour of the uniform random graphs $G_{\boldsymbol{d}}$ with a given degree sequence $\boldsymbol{d} : V \to \mathbb{N}$ is sufficiently well known. More precisely, as follows from [MR95], there exists a single parameter $Q = Q(\boldsymbol{d})$ (see (3.6) below) such that $Q < 0$ implies that $G_{\boldsymbol{d}}$ is typically sub-critical (i.e. has only small components), and $Q > 0$ implies that $G_{\boldsymbol{d}}$ is supercritical (i.e. has a giant component). Moreover, the recent paper [HM10] establishes the existence of an intermediate regime (the so-called critical window) when $Q(\boldsymbol{d})$ converges to zero at a certain rate as n tends to infinity, see also [JL09]. The results mentioned in this paragraph that will be useful in this paper are summarised in Theorem 3.2 below.

The principal contribution of this paper is thus to obtain sufficiently sharp estimates on the random degree sequence of the vacant set $\mathcal{V}^u$, and consequently on the value of the parameter Q, see Theorems 5.1, 5.2 and (6.3) below. Weaker estimates of this type were shown in [CF10], which combined with [MR95], allowed them to deduce (1.3), (1.4).

We should remark that [CF10] contains also a statement on the critical behaviour. More precisely, Theorem 2(iii) of [CF10] states that for some $u_n = u_\star(1+o(1))$ (which might be random), the size of $\mathcal{C}^{u_n}_{\max}$ is $n^{2/3+o(1)}$, $\mathbb{P}_{n,d}$-a.a.s. Our results considerably improve this statement.

Note also that much more is known about the random graphs with a given degree sequence, see for instance the results in [FR09] and [JL09]. Often the hypothesis of these results can be shown to hold true for $\mathcal{V}^u$, using Theorems 5.1 and 5.2. If this is the case, their conclusions will also apply to $\mathcal{V}^u$ ($\mathbf{P}_{n,d}$-a.a.s), providing us with more information on the geometry of the vacant set.

As an example of such application, we obtain the following improvement on the statement (1.3) about the super-critical behaviour of the vacant set.

Theorem 1.3. *Let $u < u_\star$. Then there is $\rho = \rho(u,d) \in (0,1)$ such that for every $\varepsilon > 0$*

$$n^{-1}|\mathcal{C}^{u}_{\max}| \in (\rho-\varepsilon, \rho+\varepsilon) \qquad \mathbf{P}_{n,d}\text{-a.a.s.} \tag{1.12}$$

In the above statement, the value of ρ can be explicitly calculated, see (6.9) below. Remark also that to obtain the above theorem, our precise estimates on Q are not necessary, in fact the precision obtained in [CF10] would have been sufficient.

Finally, let us briefly describe Theorems 5.1 and 5.2. The former, establishes an estimate on the expected degree distribution of $\mathcal{V}^u$, by approximating the probability that a random walk visits a neighbourhood of a given vertex $x \in V$ before time un. For this, we make use of the well known relation between random walks on graphs and discrete potential theory, as well as the pairing construction introduced by Bollobás, which we detail in Section 2. Then in Theorem 5.2 we prove that with high probability the degree sequence of $\mathcal{V}^u$ concentrates around its expectation. This is done using a standard concentration inequality, together with the fast mixing properties of the random walk on a random regular graph.

This paper is organised as follows. In Section 2 we introduce some of the notation needed in the paper and the pairing construction of random regular graphs. In Section 3, we recall the results of [CF10, HM10, JL09] needed later. Section 4 contains precise estimates on the behaviour of the simple random walk on random regular graphs. In Section 5, we give the estimates on the degree sequence of the vacant set. Theorems 1.1–1.3 are proved in Section 6. The Appendix summarises some general facts concerning random walks on finite graphs.

2. Notation and definitions

2.1. **Basic notation.** We now introduce some basic notation. Throughout the text c or c' denote strictly positive constants only depending on d, with value changing from place to place. Dependence of constants on additional parameters appears in the notation. For instance c_u denotes a positive constant depending on u and possibly on d. We write $\mathbb{N} = \{0, 1, \dots\}$ for the set of natural numbers, and $[d]$ for the set $\{1, \dots, d\}$. For a set A we denote by $|A|$ its cardinality. For any sequence of probability measures P_n and events A_n we say that A_n holds P_n-a.a.s. (asymptotically almost surely), when $\lim_{n\to\infty} P_n[A_n] = 1$.

In this paper, the term *graph* stands for a finite simple graph, that is a graph without loops or multiple edges. Sometimes we intentionally allow the graph to have loops and/or multiple edges and in this case we use the term *multigraph*. For arbitrary (multi)graph $G = (V, \mathcal{E})$, we use $\mathrm{dist}(\cdot,\cdot)$ to denote the usual graph distance and write $B(x, r)$ for the closed ball centred at x with radius r, that is $B(x, r) = \{y \in V : \mathrm{dist}(x, y) \le r\}$.

We use $\mathcal{G}_{n,d}$ (resp. $\mathcal{M}_{n,d}$) to denote the set of all d-regular graphs (resp. multigraphs) with vertex set $V_n = \{1, \dots, n\}$. Given a degree sequence $\boldsymbol{d} : V_n \to \mathbb{N}$, we use $\mathcal{G}_{\boldsymbol{d}}$ to denote the set of graphs for which every vertex $x \in V_n$ has the degree $\boldsymbol{d}_x = \boldsymbol{d}(x)$. Similarly, $\mathcal{M}_{\boldsymbol{d}}$ stands for the set of such multigraphs; here loops are counted twice when considering the degree. $\mathbb{P}_{n,d}$ and $\mathbb{P}_{\boldsymbol{d}}$ denote the uniform distributions on $\mathcal{G}_{n,d}$ and $\mathcal{G}_{\boldsymbol{d}}$ respectively.

2.2. **Pairing construction.** In order to study properties of random regular graphs, Bollobás (see e.g. [Bol01]) introduced the so-called pairing construction, which allows to generate such graphs starting from a random pairing of a set with dn elements. The same construction can be used to generate a random graph chosen uniformly at random from $\mathcal{G}_{\boldsymbol{d}}$. Since this pairing construction will be important in what follows, we give here a short overview of it.

From now on, whenever we consider a sequence $\boldsymbol{d} : V_n \to \mathbb{N}$, we suppose that $\sum_{x\in V_n} \boldsymbol{d}_x$ is even. Given such a sequence, we associate to every vertex $x \in V_n$, $\boldsymbol{d}_x$ half-edges. The set of half-edges is denoted by $H_{\boldsymbol{d}} = \{(x, i) : x \in V_n, i \in [\boldsymbol{d}_x]\}$. We write $H_{n,d}$ for the case $\boldsymbol{d}_x = d$ for all $x \in V_n$. Every perfect matching M of $H_{\boldsymbol{d}}$ (i.e. partitioning of $H_{\boldsymbol{d}}$ into $|H_{\boldsymbol{d}}|/2$ disjoint pairs) corresponds to a multigraph $G_M = (V_n, \mathcal{E}_M) \in \mathcal{M}_{\boldsymbol{d}}$ with

$$\mathcal{E}_M = \big\{\{x,y\} : \big\{(x,i),(y,j)\big\} \in M \text{ for some } i \in [\boldsymbol{d}_x],\, j \in [\boldsymbol{d}_y]\big\}. \tag{2.1}$$

We say that the matching M is simple, if the corresponding multigraph G_M is simple, that is G_M is a graph. With a slight abuse of notation, we write $\bar{\mathbb{P}}_{\boldsymbol{d}}$ for the uniform distribution on the set of all perfect matchings of $H_{\boldsymbol{d}}$, and also for the induced distribution on the set of multigraphs $\mathcal{M}_{\boldsymbol{d}}$. It is well known (see e.g. [Bol01] or [McD98]) that a $\bar{\mathbb{P}}_{\boldsymbol{d}}$ distributed multigraph G conditioned on being simple has distribution $\mathbb{P}_{\boldsymbol{d}}$, that is

$$\bar{\mathbb{P}}_{\boldsymbol{d}}[G \in \cdot\, | G \in \mathcal{G}_{\boldsymbol{d}}] = \mathbb{P}_{\boldsymbol{d}}[G \in \cdot\,], \tag{2.2}$$

and that, for d constant, there is $c > 0$ such that for all n large enough

$$c < \bar{\mathbb{P}}_{n,d}[G \in \mathcal{G}_{n,d}] < 1 - c. \tag{2.3}$$

These two claims allow to deduce $\mathbb{P}_{n,d}$-a.a.s. statements directly from $\bar{\mathbb{P}}_{n,d}$-a.a.s. statements.

The main advantage of dealing with matchings is that they can be constructed sequentially: To construct a uniformly distributed perfect matching of $H_{\boldsymbol{d}}$ one samples *without replacements* a sequence $h_1, \dots, h_{|H_{\boldsymbol{d}}|}$ of elements of $H_{\boldsymbol{d}}$ in the following way. For i odd, h_i can be chosen by an arbitrary rule (which might also depend on the previous $(h_j)_{j<i}$), while if i is even, h_i must be chosen uniformly among the remaining half-edges. Then, for every $1 \le i \le |H_{\boldsymbol{d}}|/2$ one matches h_{2i} with h_{2i-1}.

It is clear from the above construction that, conditionally on $M' \subseteq M$ for a (partial) matching M' of $H_{\boldsymbol{d}}$, $M \setminus M'$ is distributed as a uniform perfect matching of $H_{\boldsymbol{d}} \setminus \{(x,i) : (x,i) \text{ is matched in } M'\}$. Since the law of the graph G_M does not depend on the labels 'i' of the half-edges, we obtain for all partial matchings M' of $H_{\boldsymbol{d}}$

$$\bar{\mathbb{P}}_{\boldsymbol{d}}[G_{M\setminus M'} \in \cdot\, | M \supset M'] = \bar{\mathbb{P}}_{\boldsymbol{d}'}[G_M \in \cdot], \tag{2.4}$$

where $\boldsymbol{d}'_x$ is the number of half-edges incident to x in $H_{\boldsymbol{d}}$ that are not yet matched in M', that is $\boldsymbol{d}'_x = \boldsymbol{d}_x - \big|\{\{(y_1,i),(y_2,j)\} \in M' : y_1 = x, i \in [\boldsymbol{d}_x]\}\big|$, and $G_{M\setminus M'}$ is the graph corresponding to a non-perfect matching $M \setminus M'$, defined in the obvious way.

2.3. **Random walk notation.** For an arbitrary multigraph $G = (V, \mathcal{E})$, we use P_x^G to denote the law of canonical continuous-time simple random walk on G started at $x \in V$, that is of the Markov process with generator given by

$$\mathcal{L}f(x) = \sum_{y\in V} p_{xy}(f(y) - f(x)), \qquad \text{for } f : V \to \mathbb{R}, x \in V. \tag{2.5}$$

Here $p_{xy} = n_{xy}/d_x$, n_{xy} is the number of edges connecting x and y in G, and d_x is the degree of x; the loops are counted twice in n_{xx} and d_x.

We write $P_x^{G,\ell}$ for the restriction of P_x^G to $D([0,\ell], V)$ and $P_{xy}^{G,\ell}$ for the law of random walk bridge, that is for $P_x^{G,\ell}$ conditioned on $X_\ell = y$. We write $E_x^G, E_x^{G,\ell}, E_{xy}^{G,\ell}$ for the corresponding expectations. The canonical shifts on $D([0,\infty), V)$ are denoted by θ_t. The time of the n-th jump is denoted by τ_n, i.e. $\tau_0 = 0$ and for $n \ge 1$, $\tau_n = \inf\{t \ge 0 : X_t \ne X_0\} \circ \theta_{\tau_{n-1}} + \tau_{n-1}$. The process counting the number of jumps before time t is denoted by $N_t = \sup\{k : \tau_k \le t\}$. Note that, when G is simple, under P_x^G, $(N_t)_{t\ge 0}$ is a Poisson process on $\mathbb{R}_+$ with intensity one. We write $\hat{X}_n$, $n \in \mathbb{Z}_+$, for the discrete skeleton of the process X_t, that is $\hat{X}_n = X_{\tau_n}$.

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

$$H_A = \inf\{t \ge 0 : X_t \in A\}, \qquad \text{and} \qquad \tilde{H}_A = H_A \circ \theta_{\tau_1} + \tau_1. \tag{2.6}$$

We denote by π the stationary distribution for the simple random walk on G, which is uniform if G is d-regular (even if G is not simple). P^G stands for the law of the simple random walk started at π and E^G for the corresponding expectation. For all real valued functions f, g on V we define the Dirichlet form

$$\mathcal{D}(f,g) = \frac{1}{2} \sum_{x,y \in V} (f(x) - f(y))(g(x) - g(y))\pi_x p_{xy} = -\sum_{x \in G} \mathcal{L}f(x)g(x)\pi_x. \tag{2.7}$$

The spectral gap of G is given by

$$\lambda_G = \min\big\{\mathcal{D}(f,f) : \pi(f^2) = 1, \pi(f) = 0\big\}. \tag{2.8}$$

From [SC97], p. 328, it follows that for d-regular graphs,

$$\sup_{x,y \in V} |P_x[X_t = y] - \pi_y| \le e^{-\lambda_G t}, \text{ for all } t \ge 0. \tag{2.9}$$

It is also a well known fact (see e.g. [Fri08]) that there exist $\alpha > 0$ such that

$$\lambda_G > \alpha, \qquad \text{both } \mathbb{P}_{n,d}\text{-a.a.s. and } \bar{\mathbb{P}}_{n,d}\text{-a.a.s.} \tag{2.10}$$

3. Preliminaries

3.1. **Distribution of the vacant set.** Recall the notation $\mathcal{V}^u = \mathcal{V}^u_G$ for the vacant set of the random walk on the graph $G = (V, \mathcal{E})$ at level u, (1.1). For the purpose of this paper, it is suitable to define a closely related object, the *vacant graph* $\mathbf{V}^u = (V, \mathcal{E}^u)$ where

$$\mathcal{E}^u = \{\{x,y\} \in \mathcal{E} : x, y \in \mathcal{V}^u_G\}. \tag{3.1}$$

It is important to notice that the vertex set of $\mathbf{V}^u$ is V and not $\mathcal{V}^u$, in particular $\mathbf{V}^u$ is not the graph induced by $\mathcal{V}^u$ in G. Observe however that the maximal connected component of the vacant set $\mathcal{C}_{\max}$ (defined before in terms of the graph induced by $\mathcal{V}^u$ in G) coincides with the maximal connected component of the vacant graph $\mathbf{V}^u$ (except when $\mathcal{V}^u$ is empty, but this difference can be ignored in our investigations).

We use $\mathcal{D}^u : V \to \mathbb{N}$ to denote the (random) degree sequence of $\mathbf{V}^u$, and write $Q^u_{n,d}$ for the distribution of this sequence under the annealed measure $\bar{\mathbf{P}}_{n,d}$, defined by $\bar{\mathbf{P}}_{n,d}(\cdot) := \int P^G(\cdot)\bar{\mathbb{P}}_{n,d}(\mathrm{d}G)$.

The following proposition from Cooper and Frieze [CF10] allows us to reduce questions on the properties of the vacant set $\mathcal{V}^u$ of the random walk on random regular graphs to questions on random graphs with given degree sequences.

Proposition 3.1 (Lemma 6 of [CF10])**.** *For every $u \ge 0$, the distribution of the vacant graph $\mathbf{V}^u$ under $\bar{\mathbf{P}}_{n,d}$ is given by $\bar{\mathbb{P}}_{\boldsymbol{d}}$ where $\boldsymbol{d}$ is sampled according to $Q^u_{n,d}$, that is*

$$\bar{\mathbf{P}}_{n,d}[\mathbf{V}^u \in \cdot\,] = \int \bar{\mathbb{P}}_{\boldsymbol{d}}[G \in \cdot\,]Q^u_{n,d}(\mathrm{d}\boldsymbol{d}). \tag{3.2}$$

Although a proof of Lemma 3.1 can be found in [CF10], we provide a proof here for the sake of completeness.

Proof. Let M be a $\bar{\mathbb{P}}_{n,d}$-distributed pairing of $H_{n,d}$ and let X be a random walk on $G = G_M$. Define $M_t \subset M$ to be the set of all pairs of half-edges incident to a vertex X_s with $s \le t$,

$$M_t = \big\{\{(x,i),(y,j)\} \in M : x \in \{X_s : s \le t\}, i \in [d]\big\}. \tag{3.3}$$

It is easy to see that the edges of the vacant graph $\mathbf{V}^u$ correspond exactly to the pairs in $M \setminus M_{un}$, that is $\mathbf{V}^u = G_{M \setminus M_{un}}$. In particular, $\mathcal{D}^u(x)$ is the number of the half-edges incident to x not matched in M_{un}. Denoting by $\mathcal{F}_u$ the σ-algebra generated by $((X_s, M_s), s \le un)$, the above implies that $\mathcal{D}^u$ is $\mathcal{F}_u$-measurable.

It follows from (2.4) that, conditionally on $\mathcal{F}_u$, the distribution of $G_{M\setminus M_{un}}$ only depends on the sequence of half-edges that are not matched in M_{un}, and is given by $\bar{\mathbb{P}}_{\mathcal{D}_u}$. More precisely,

$$\bar{\mathbf{P}}_{n,d}[\mathbf{V}^u \in \cdot\,|\mathcal{F}_u] = \bar{\mathbf{P}}_{n,d}[G_{M\setminus M_{un}} \in \cdot\,|\mathcal{D}^u] = \bar{\mathbb{P}}_{\mathcal{D}^u}[G \in \cdot\,], \tag{3.4}$$

and thus

$$\bar{\mathbf{P}}_{n,d}[\mathbf{V}^u \in \cdot\,] = \bar{\mathbf{P}}_{n,d}\big[\bar{\mathbf{P}}_{n,d}[\mathbf{V}^u \in \cdot\,|\mathcal{F}_u]\big] = \bar{\mathbf{P}}_{n,d}\big[\bar{\mathbb{P}}_{\mathcal{D}^u}[G \in \cdot\,]\big] = \int \bar{\mathbb{P}}_{\boldsymbol{d}}[G \in \cdot\,]Q^u_{n,d}(\mathrm{d}\boldsymbol{d}), \tag{3.5}$$

where the last equality follows from the definition of $Q^u_{n,d}$. This concludes the proof of Proposition 3.1. □

3.2. **Behaviour of random graphs with a given degree sequence.** We now summarise the results about the behaviour of random graphs with a given degree sequence which will be used in this paper. For a degree sequence $\boldsymbol{d} : V_n \to \mathbb{N}$ we define

$$Q(\boldsymbol{d}) = \frac{\sum_{x=1}^n \boldsymbol{d}_x^2}{\sum_{x=1}^n \boldsymbol{d}_x} - 2, \tag{3.6}$$

and set $n_i(\boldsymbol{d})$ to be the number of $x \le n$ with $\boldsymbol{d}_x = i$,

$$n_i(\boldsymbol{d}) = \big|\{x \in V_n : \boldsymbol{d}_x = i\}\big|. \tag{3.7}$$

For any graph G we use $\mathcal{C}_{\max}(G)$ to denote the maximal connected component of G.

The following theorem summarises the results of [MR95, JL09, HM10] needed later.

Theorem 3.2. *Let $(\boldsymbol{d}^n)_{n\ge1}$, $\boldsymbol{d}^n : V_n \to \mathbb{N}$, be a sequence of degree sequences. We assume that the degrees are uniformly bounded (i.e. $\max\{\boldsymbol{d}^n_x : n \ge 1, x \le n\} \le \Delta < \infty$), and that $n_1(\boldsymbol{d}^n) \ge \zeta n$ for a $\zeta > 0$.*

(i) *(critical window) If $|Q(\boldsymbol{d}^n)| \le \lambda n^{-1/3}$ for all $n \ge 1$, then for every $\varepsilon > 0$ there exists $A = A(\zeta, \lambda, \varepsilon, \Delta)$ such that for all n large enough*

$$\bar{\mathbb{P}}_{\boldsymbol{d}^n}[A^{-1}n^{2/3} \le |\mathcal{C}_{\max}(G)| \le An^{2/3}] \ge 1 - \varepsilon.$$

(ii) *(below the window) If $\lim_{n\to\infty} n^{1/3}Q(\boldsymbol{d}^n) = -\infty$ and $\lim_{n\to\infty} Q(\boldsymbol{d}^n) = 0$, then for every $\varepsilon > 0$ exists $B = B(\zeta, \varepsilon, \Delta) < \infty$ such that for all n large enough*

$$\bar{\mathbb{P}}_{\boldsymbol{d}^n}\Big[|\mathcal{C}_{\max}(G)| < B\sqrt{n/|Q(\boldsymbol{d}^n)|}\Big] > 1 - \varepsilon.$$

(iii) *(above the window) Let $\lim_{n\to\infty} n^{1/3}Q(\boldsymbol{d}^n) = +\infty$ and $\lim_{n\to\infty} Q(\boldsymbol{d}^n) = 0$. In addition, assume that*

$$\lim_{n\to\infty} \frac{n_i(\boldsymbol{d}^n)}{n} = p_i, \qquad \text{for all } 0 \le i \le \Delta, \tag{3.8}$$

for some probability distribution $(p_i)_{0\le i\le\Delta}$ on $\{0, \dots, \Delta\}$, and set $\lambda = \sum_{i=0}^{\Delta} ip_i$, $\beta = \sum_{i=0}^{\Delta} i(i-1)(i-2)p_i$, and $v_n = 2n\lambda^2\beta^{-1}Q(\boldsymbol{d}^n)$. Then, for every ε

$$\Big|\frac{|\mathcal{C}_{\max}(G)|}{v_n} - 1\Big| < \varepsilon, \qquad \bar{\mathbb{P}}_{\boldsymbol{d}^n}\text{-a.a.s.},$$

(iv) *(super-critical regime) Let $\lim_{n\to\infty} Q(\boldsymbol{d}^n) = Q_\infty > 0$ and assume that (3.8) holds. Let g be the generating function of (p_i), $g(x) = \sum_{i=0}^{\Delta} p_i x^i$. Then there exists a unique solution ξ to $g'(x) = \lambda x$ in $(0, 1)$, and for $\rho = 1 - g(\xi)$ and any $\varepsilon > 0$*

$$\Big|\frac{|\mathcal{C}_{\max}(G)|}{n} - \rho\Big| \le \varepsilon, \qquad \bar{\mathbb{P}}_{\boldsymbol{d}^n}\text{-a.a.s.}$$

Proof. Parts (i), (ii) correspond to Theorems 1.1 and 1.2 of [HM10], where these statements are proved under more general assumptions. In particular, [HM10] does not require the uniform upper bound Δ on the maximal degree. The restriction to the uniformly bounded degree sequences implies that the constant $R(\boldsymbol{d}^n)$ used in [HM10] satisfies $c < R(\boldsymbol{d}^n) < c^{-1}$ for all n large enough and is therefore immaterial for our purposes.

Parts (iii), (iv) are taken from Theorems 2.3 and 2.4 of [JL09]. When reading those theorems it is useful to realise that the (p_i)-distributed random variable D used in [JL09] satisfies $\mathbb{E}[D] = \lambda$ and that $E[D(D-2)] = \lambda Q_\infty$ in our notation.

Remark however that neither [HM10], or [JL09] consider degree sequences with vertices of degree zero, that is with $n_0(\boldsymbol{d}^n) > 0$. It can however be seen easily, that if $n_0(\boldsymbol{d}^n)$ does not exceed $\zeta' n$, $\zeta' < 1$ (which is implied by the assumptions of the theorem), the vertices of degree zero do not have any influence on the existence of the giant cluster, they only change the constants A, B in (i), (ii). For (iii), (iv), when $n_0(\boldsymbol{d}^n)/n \to p_0 \neq 0$, one applies the theorem for the modified sequence $(\bar{\boldsymbol{d}}^n)$ where all vertices of degree 0 are omitted. The new degree sequences $\bar{\boldsymbol{d}}^n$ are functions on $V_{\bar{n}}$, with $\bar{n} = n(1-p_0) + o(1)$. They satisfy $n_i(\bar{\boldsymbol{d}}^n)/\bar{n} \xrightarrow{n\to\infty} p_i/(1-p_0) =: \bar{p}_i$. Therefore, denoting by the letters with bars the quantities related to the distribution $(\bar{p}_i)$, we obtain $Q(\bar{\boldsymbol{d}}^n) = Q(\boldsymbol{d}^n)$, $\bar{\lambda} = \lambda/(1-p_0)$, $\bar{\beta} = \beta/(1-p_0)$, $\bar{g}(x) = (g(x) - p_0)/(1-p_0)$, and $\bar{\xi} = \xi$. This implies that $\bar{v}_{\bar{n}} = v_n$ and $\bar{\rho}\bar{n} = \rho n$, confirming that zero-degree vertices have no influence on the asymptotics of the size of the maximal connected component. This completes the proof. □

4. Random walk estimates

This section contains estimates on the random walk on random regular multigraphs which will be useful later in order to estimate the typical degree sequence of the vacant graph $\mathbf{V}^u$.

We start by introducing some notation. We use $\mathbb{T}^d$ to denote the infinite d-regular tree with root $\varnothing$. For a d-regular multigraph $G = (V, \mathcal{E})$, a map ϕ from $\mathbb{T}^d \to V$ is said to be a *covering of G from $x \in V$*, if $\phi(\varnothing) = x$, and for every $y \in \mathbb{T}^d$, ϕ maps the d neighbours of y in $\mathbb{T}^d$ to the neighbours of $\phi(y)$ in G, including the multiplicities and the loops. For d-regular multigraphs constructed by the pairing construction this means that the neighbours of y are sent by ϕ to the vertices which are paired with $(\phi(y), i)$, $i \in [d]$.

In agreement with our previous notations, $P_y^{\mathbb{T}^d}$ denotes the law of the continuous-time simple random walk on $\mathbb{T}^d$ starting from y. It is important to notice that fixing a covering ϕ from x, the image by ϕ of a random walk in $\mathbb{T}^d$ with law $P_\varnothing^{\mathbb{T}^d}$ is distributed as P_x^G.

For every finite connected $\mathbb{A} \subset \mathbb{T}^d$ and $z \in \mathbb{A}$ we now define the escape probabilities

$$e_{\mathbb{A}}(z) = P_z^{\mathbb{T}^d}[\tilde{H}_{\mathbb{A}} = \infty], \tag{4.1}$$

which can be calculated explicitly in practical examples using the fact that

$$P_y^{\mathbb{T}^d}[H_\varnothing = \infty] = \frac{d-2}{d-1}, \tag{4.2}$$

for every neighbour y of $\varnothing$. This comes from a standard calculation for a one-dimensional simple random walk with drift, see for instance [Woe00], proof of Lemma (1.24).

We use $z_i \in \mathbb{T}^d$, $i \in [d]$ to denote the neighbours of $\varnothing$ listed in some predefined order. For $x \in \mathcal{V}$, let ϕ_x be a covering of G from x. From now on, if G was obtained by the pairing construction, we require that $\phi_x(z_i)$ is the vertex matched with (x, i). Otherwise ϕ_x can be chosen arbitrarily, since our statements will not depend on which particular choice of ϕ_x is picked. For any $D \subset [d]$, we define the sets

$$\mathbb{B}_D = \{\varnothing\} \cup \{z_i : i \in D\} \quad \text{and} \quad B_{x,D} = \{x\} \cup \{\phi_x(z_i) : i \in D\}. \tag{4.3}$$

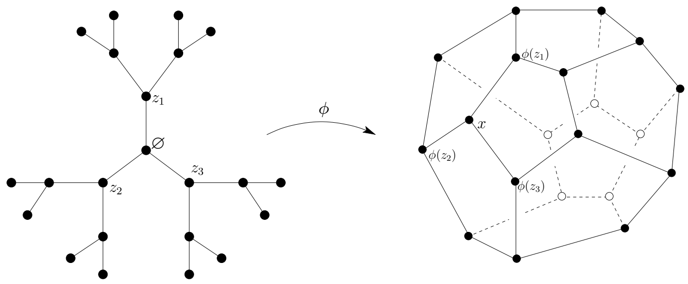

FIGURE 1. A covering ϕ of a regular graph from x.

The sets $B_{x,D}$ will be used later in the calculation of the degree distribution of the vacant set: using an inclusion-exclusion formula, we can express the event {the degree of x in $\mathbf{V}^u$ is k} in terms of events $\{B_{x,D} \subset \mathcal{V}^u\}$, for $D \subset [d]$.

We first prove a technical lemma describing the graph V after removing the set $B_{x,D}$.

Lemma 4.1. *For $K > 0$, we say that the graph G is K-*good*, if there is no $x \in V$ and $D \subset [d]$ such that:*

(a) *$B(x, 3)$ is a tree, and*
(b) *either $V \setminus B_{x,D}$ is disconnected or* $\mathrm{diam}(V \setminus B_{x,D}) \geq K \log n$.

Then, there is $K > 1$ such that G is K-good, $\bar{\mathbb{P}}_{n,d}$-a.a.s.

Proof. To prove this claim we use Lemma 2.14 of [Wor99] which states that $\bar{\mathbb{P}}_{n,d}$-probability that there exists a (s, j)-separating set in G, that is a set $S \subset V$ with $|S| = s$ such that $G \setminus S$ contains a component of exactly j vertices, satisfies

$$\bar{\mathbb{P}}_{n,d}[G \text{ has an } (s, j)\text{-separating set}] \leq 3^{2+s/d} \Big(\frac{j+s}{n}\Big)^{j(\frac{d}{2}-1)} n^{\frac{s}{2}} (j+s)^{\frac{3}{2}s}. \tag{4.4}$$

(This formula is used in [Wor99] to prove that G is $\mathbb{P}_{n,d}$-a.a.s. d-connected. However, the d-connectedness cannot be used directly to prove our claim, since $|B(x, 1)| = d + 1$.)

Observe first that it is sufficient to consider $D = [d]$. Indeed, if x is such that $B(x, 3)$ is a tree and the graph is connected after removing $B_{x,[d]}$, then it is connected after removing $B_{x,D}$ for any $D \subset [d]$. In addition, obviously $\mathrm{diam}(G \setminus B_{x,D}) \leq \mathrm{diam}(G \setminus B_{x,[d]}) + 2$.

Let now $D = [d]$, that is $B_{x,D} = B(x, 1)$. Assume that $G \setminus B(x, 1)$ is disconnected. Then removing the set $B(x, 1) \setminus \{x\}$ from G, divides the graph into at least three components. One of them is $\{x\}$, and the other ones are are contained in $G \setminus B(x, 1)$. Since we require that $B(x, 3)$ is a tree and since G is d-regular, the size of these other components is at least 4. We thus apply (4.4) with $s = d$ and $j \geq 4$: The $\bar{\mathbb{P}}_{n,d}$-probability that there is $x \in V$ such that $B(x, 3)$ is a tree and $G \setminus B(x, 1)$ is disconnected is bounded from above by

$$\sum_{j=4}^{n/2} \bar{\mathbb{P}}_{n,d}[G \text{ has an } (s, j)\text{-separating set}]. \tag{4.5}$$

The largest term in this sum, corresponding to $j = 4$, is $O(n^{4-\frac{3}{2}d}) = o(1)$. All the remaining terms are much smaller. Actually, as in the proof of Theorem 2.10 of [Wor99], it can be shown that (4.5) tends to 0 as $n \to \infty$. Hence, $\bar{\mathbb{P}}_{n,d}$-a.a.s. there is no $x \in V$ such that $B(x, 3)$ is a tree and $G \setminus B(x, 1)$ is disconnected.

To bound the diameter of $G \setminus B(0, 1)$ we use the fact that $\bar{\mathbb{P}}_{n,d}$-a.a.s. $\mathrm{diam}(G) \leq 2 \log_{d-1} n$, see [BF82]. We may also assume that $G \setminus B(x, 1)$ is connected, which as we proved occurs $\bar{\mathbb{P}}_{n,d}$-a.a.s.

We now claim that

$$\text{removing one vertex } v \text{ of degree } d \text{ in an arbitrary graph while keeping it connected can increase the diameter of the graph at most by a factor of } 3^{d-1}. \tag{4.6}$$

To prove this claim we first consider the removal of an edge: Removing one edge e from a graph while keeping in connected can increase the diameter of the graph at most by factor 3. To see this it is sufficient to consider the shortest path μ in $G \setminus \{e\}$ connecting the vertices of e (such path must exist since $G \setminus \{e\}$ is connected, and cannot be longer than $2 \operatorname{diam} G$), and to replace the edge e by the path μ in every geodesic of G that contains e.

Having understood the removal of edges, we can analyse the removal of a vertex v. We first remove all but one of the edges of G incident to v, in this procedure the diameter of the graph is multiplied by at most 3^{d-1}. Removing v together with the last edge linking it to $G \setminus \{v\}$ yields the claim (4.6).

The claim (4.6) and $|B(x,1)| \le d+1$ then imply that $\bar{\mathbb{P}}_{n,d}$-a.a.s

$$\operatorname{diam}(G \setminus B(x,1)) \le 3^{(d-1)(d+1)} \operatorname{diam} G \le 3^{(d-1)(d+1)} \cdot 2 \log_{d-1} n. \tag{4.7}$$

This completes the proof of the lemma. □

We now start controlling how the random walk visits the sets $B_{x,D}$ defined in (4.3). In the lemma below, we show that the probability of escaping from $B_{x,D}$ for a large time can be approximated by the escape probability on the infinite tree, defined in (4.1).

Lemma 4.2. *For every $x \in V_n$, $D \subset [d]$ and $i \in D$*

$$\bar{\mathbb{E}}_{n,d}\Big[\Big|P^G_{\phi_x(z_i)}[\tilde{H}_{B_{x,D}} > \log^2 n] - e_{\mathbb{B}_D}(z_i)\Big|\Big] \le \frac{c \log^4 n}{n}. \tag{4.8}$$

and

$$\bar{\mathbf{P}}_{n,d}[\tilde{H}_{B_{x,D}} \le \log^2 n] \le \frac{c \log^2 n}{n}. \tag{4.9}$$

Proof. To simplify the notation we write B, $\mathbb{B}$, z and $\phi(z)$ for B_D, $\mathbb{B}_{x,D}$, z_i and $\phi_x(z_i)$. Using the fact that ϕ_x maps a random walk on $\mathbb{T}^d$ to a random walk on G, we can write

$$\begin{aligned} P^G_{\phi(z)}[\tilde{H}_B \le \log^2 n] &= P^{\mathbb{T}^d}_z[\tilde{H}_{\phi_x^{-1}(B)} \le \log^2 n] \\ &= P^{\mathbb{T}^d}_z[\tilde{H}_{\mathbb{B}} \le \log^2 n] + P^{\mathbb{T}^d}_z[\tilde{H}_{\phi_x^{-1}(B)\setminus\mathbb{B}} \le \log^2 n, \tilde{H}_{\mathbb{B}} > \log^2 n]. \end{aligned} \tag{4.10}$$

Therefore, the left-hand side of (4.8) can be bounded from above by

$$\Big|P^{\mathbb{T}^d}_z[\tilde{H}_{\mathbb{B}} > \log^2 n] - e_{\mathbb{B}}(z)\Big| + \bar{\mathbb{E}}_{n,d}\Big[P^{\mathbb{T}^d}_z[\tilde{H}_{\phi_x^{-1}(B)\setminus\mathbb{B}} \le \log^2 n]\Big]. \tag{4.11}$$

Using the Markov property at time $\log^2 n$ and the definition of $e_{\mathbb{B}}(z)$, the first term equals

$$\begin{aligned} &P^{\mathbb{T}^d}_z[\log^2 n < \tilde{H}_{\mathbb{B}} < \infty] \\ &\le P^{\mathbb{T}^d}_z\Big[d(\varnothing, X_{\log^2 n}) \le \frac{d-2}{2d}\log^2 n\Big] + \sup_{u: d(u,\varnothing) > \frac{d-2}{2d}\log^2 n} P^{\mathbb{T}^d}_u\Big[H_\varnothing < \infty\Big]. \end{aligned} \tag{4.12}$$

Since $d(X_t, \varnothing)$ under $P^{\mathbb{T}^d}_z$ is a random walk on $\mathbb{N}$ with expected drift given by $(d-2)/d$, both terms above are bounded by $c \exp\{-c' \log^2 n\}$.

To bound the second term in (4.11), note that if at time t the random walk on $\mathbb{T}_d$ started from z visits a point in $\phi_x^{-1}(B_{x,D}) \setminus \mathbb{B}$, then the trajectory of the image walk on G together with $B(\phi(z), 2)$ contains a cycle in G. Therefore, denoting by $G|_A$ the subgraph of G generated by $A \subset V_n$,

$$\bar{\mathbb{E}}_{n,d}\Big[P_z^{\mathbb{T}^d}[\tilde{H}_{\phi_x^{-1}(B)\setminus\mathbb{B}} \le \log^2 n]\Big] \le \bar{\mathbf{P}}_{n,d}\Big[G|_{B(X_0,2)\cup\{X_s : s \le \log^2 n\}} \text{ contains a cycle}\Big]. \tag{4.13}$$

Taking care of the possibility that the continuous-time random walk makes more than $2\log^2 n$ steps before time $\log^2 n$, using the notation from Section 2.3, this is bounded from above by

$$P_z^{\mathbb{T}^d}[N_{\log^2 n} \ge 2\log^2 n] + \bar{\mathbf{P}}_{n,d}\Big[G|_{B(X_0,2)\cup\{\hat{X}_i : i \le 2\log^2 n\}} \text{ contains a cycle}\Big] \tag{4.14}$$

The random variable $N_{\log^2 n}$ has Poisson distribution with mean $\log^2 n$, therefore the first term in (4.14) is smaller than $ce^{-c'\log^2 n}$. To bound the second term, observe that the considered subgraph can be constructed inductively by the following variant of the construction from Section 2.2:

(1) Choose $\hat{X}_0$ uniformly at random in V_n. Use the pairing construction of Section 2.2 to construct the set $B(\hat{X}_0, 2)$. This requires creating at most $d + d(d-1)$ pairs.
(2) Let $\hat{X}_1$ be a uniformly chosen neighbour of $\hat{X}_0$. (All neighbours of X_0 are known from the first step.)
(3) Repeat for all $i \in \{2, \dots, 2\log^2 n\}$ the following steps:
   (a) Choose Z_i uniformly in $[d]$, independently of the previous randomness.
   (b) If the half-edge $(\hat{X}_{i-1}, Z_i)$ is not yet matched, then match it with an half-edge chosen uniformly among the remaining half-edges, as in the pairing construction.
   (c) Let $\hat{X}_i$ be the vertex that is matched with the half-edge $(\hat{X}_{i-1}, Z_i)$.

The probability that a cycle is created in the step (1) is easily bounded by c/n. It is not possible to create any cycle in the step (2). In the step (3) the cycle is created only when the half-edge $(\hat{X}_{i-1}, Z_i)$ is not yet matched (otherwise we do not add any new edge to the subgraph) and when the vertex $\hat{X}_i$ was already 'visited by the algorithm', that is it has some matched half-edges from the previous steps. Since the algorithm visits at most $d + d(d-1) + 2\log^2 n$ vertices, the probability to match $(\hat{X}_{i-1}, Z_i)$ with an already visited vertex is smaller than $cn^{-1}\log^2 n$. Therefore the probability to create the cycle in the step (3) is at most $cn^{-1}\log^4 n$. In consequence, the second term in (4.11) is smaller than $ce^{-c'\log^2 n} + cn^{-1}\log^4 n$ which establishes (4.8).

In order to prove (4.9), observe that

$$\bar{\mathbf{P}}_{n,d}[H_B \le \log^2 n] \le \bar{\mathbf{P}}_{n,d}[N_{\log^2 n} \ge 2\log^2 n] + \sum_{i=0}^{2\log^2 n} \bar{\mathbf{P}}_{n,d}\Big[\hat{X}_i \in B\Big]. \tag{4.15}$$

Since under $\bar{\mathbf{P}}_{n,d}$, the random vertex $\hat{X}_i$ is uniformly distributed in V_n, $\bar{\mathbf{P}}_{n,d}[\hat{X}_i \in B] \le n^{-1}|B| \le (d+1)n^{-1}$. Hence,

$$\bar{\mathbf{P}}_{n,d}[H_B \le \log^2 n] \le c'e^{-c\log^2 n} + 2(d+1)n^{-1}\log^2 n \le cn^{-1}\log^2 n. \tag{4.16}$$

This completes the proof of Lemma 4.2. □

The previous lemma has a simple corollary.

Corollary 4.3. *Fix $D \subset [d]$ and $i \in D$, then both $\bar{\mathbb{P}}_{n,d}$-a.a.s. and $\mathbb{P}_{n,d}$-a.a.s.*

$$\Big|\Big\{x \in V_n : \big|P^G_{\phi_x(z_i)}[\tilde{H}_{B_{x,D}} > \log^2 n] - e_{\mathbb{B}_D}(z_i)\big| > n^{-1/2}\Big\}\Big| \le (\log^5 n)n^{1/2} \tag{4.17}$$

and

$$\left|\left\{x \in V_n : P^G[H_{B_{x,D}} \le \log^2 n] > n^{-1/2}\right\}\right| \le (\log^3 n) n^{1/2}. \tag{4.18}$$

Proof. Note that the complements of the above events are respectively contained in the events $\sum_{x \in V_n} |P^G_{\phi_x(z_i)}[\tilde{H}_{B_{x,D}} \le \log^2 n] - e_{\mathbb{B}}(z_i)| > \log^5 n$ and $\sum_{x\in V_n} P[H_{B_{x,D}} \le \log^2 n] > \log^3 n$. Thus, the $\bar{\mathbb{P}}_{n,d}$-a.a.s. statements of the corollary are implied by the Markov inequality and Lemma 4.2. The $\mathbb{P}_{n,d}$-a.a.s. statements then follow using the remark below (2.3). □

In what follows we will make use of the quasi-stationary distribution with respect to a set $B \subset V$. The usual definition of this distribution is given in the Appendix, see (A.1). The quasi-stationary distribution can be thought as the asymptotic distribution of the position of a random walk conditioned not to visit B. To proceed, we will need the following lemma which gives the rate of convergence of such conditioned walk towards the quasi-stationary distribution.

Lemma 4.4. *Let $t_n = \log^2 n$. Then, $\bar{\mathbb{P}}_{n,d}$-a.a.s., for any x such that $B(x, 3)$ is a tree and for any connected $A \subset B(x, 1)$*

$$\sup_{x,y \in V\setminus A} \left| P^G_x[Y_{t_n} = y | H_A > t_n] - \sigma_A(y)\right| \le c e^{-c' \log^2 n}. \tag{4.19}$$

Proof. By Lemma 4.1, $V \setminus A$ is $\bar{\mathbb{P}}_{n,d}$-a.a.s. connected and has diameter smaller than $K \log n$. To prove the claim of the lemma we are going to make use of Lemmas A.1 and A.2, which are general results on Markov chains presented in the Appendix. We use the notation introduced there, in particular $0 \le \lambda^A_1 \le \lambda^A_2 \le \dots$ stand for the eigenvalues of the $(-\mathcal{L}^A)$ where $\mathcal{L}^A$ is the generator of the random walk killed on hitting A.

From (2.10), $\lambda_G \ge \alpha > 0$, $\mathbb{P}_{n,d}$-a.a.s. Therefore, using Lemma A.1, we obtain that

$$\lambda^A_2 - \lambda^A_1 \ge \alpha - \frac{1}{E[H_A]}, \qquad \mathbb{P}_{n,d} - \text{a.a.s.}, \tag{4.20}$$

To obtain an upper bound on $E[H_A]^{-1}$, we use Proposition 3.1 and (2.10) of [ČTW11]. Using this proposition with $A = A$ and $C = V \setminus A$, we obtain

$$\frac{1}{E[H_A]} \le c\pi(A) \le \frac{c}{n}. \tag{4.21}$$

This implies that $\mathbb{P}_{n,d}$-a.a.s. $\lambda^A_2 - \lambda^A_1 \ge c > 0$.

Using the above fact together with Lemma A.2 (observing that $\pi(x) = 1/n$ on regular graphs) we obtain that $\bar{\mathbb{P}}_{n,d}$-a.a.s.

$$\sup_{x,y\in V\setminus A} \left|P^G_x[Y_{t_n} = y|H_A > t_n] - \sigma_A(y)\right| \le \frac{cn^{3/2} \exp\{-c \log^2 n\}}{\inf_{z\in V\setminus A} \sigma_A(z)}. \tag{4.22}$$

We have to bound the infimum in the denominator. For this, take $z \in V \setminus A$ and any $t \ge 0$. By reversibility, for any $z' \in V \setminus A$,

$$P^G_{z'}[X_t = z|H_A > t] = P^G_z[X_t = z'|H_A > t] \frac{P^G_z[H_A > t]}{P^G_{z'}[H_A > t]}. \tag{4.23}$$

In order to bound the above ratio, note that

$$P^G_z[H_A > t] \ge P^G_z[H_{z'} < H_A, H_A \circ \theta_{H_{z'}} > t] = P^G_z[H_{z'} < H_A] P^G_{z'}[H_A > t]. \tag{4.24}$$

As $\bar{\mathbb{P}}_{n,d}$-a.a.s. the graph induced by $V \setminus A$ has diameter at most $K \log n$, we can find a path γ with length at most $K \log n$, connecting z and z' and not passing through A. This gives us that $P_z[H_{z'} < H_A] \geq d^{-c\log n} \geq cn^{-c'}$. From Lemma A.2 $\lim_{t\to\infty} P_w[X_t = z | H_A > t] = \sigma_A(z)$ uniformly for all $w, z \in V \setminus A$. Therefore, taking the limit $t \to \infty$ in (4.23), $\sigma_A(z) \geq c\sigma_A(z')n^{-c'}$. Together with the fact that σ_A is a probability measure this yields

$$\inf_{z\in V\setminus A} \sigma_A(z) \geq cn^{-c'}. \tag{4.25}$$

Using the above result with (4.22) finishes the proof of Lemma 4.4. □

5. Degree sequence of the vacant graph

We are now in position to estimate the typical degree sequence of the vacant graph $\mathcal{V}^u_G$ under $\bar{\mathbf{P}}_{n,d}$. At this point it is instructive to mention the relation between this set and the random interlacements process on the d-regular tree $\mathbb{T}^d$.

The model of random interlacements on transient weighted graphs is constructed in [Tei09], see also [Szn10] for the original construction of the model in the particular case of $\mathbb{Z}^d$, $d \geq 3$. Random interlacement on $\mathbb{T}^d$ can be understood as a measure Q^u on the space $\{0,1\}^{\mathbb{T}^d}$ which samples a random subset $\mathcal{V}^u_{\mathbb{T}^d}$ of $\mathbb{T}^d$ (called the vacant set left by random interlacements at level u) characterised by the following:

$$Q^u[K \subset \mathcal{V}^u_{\mathbb{T}^d}] = \exp\{-u \operatorname{cap}(K)\}, \qquad \text{for every finite } K \subset \mathbb{T}^d, \tag{5.1}$$

where $\operatorname{cap}(K) = \sum_{x\in K} e_K(x)$ for e_K as in (4.1), cf. (2.27) of [Tei09].

Intuitively speaking, the vacant set of the random interlacement $\mathcal{V}^u_{\mathbb{T}^d}$ gives the asymptotic local picture of $\mathcal{V}^u_G$ under $\mathbf{P}_{n,d}$ as n tends to infinity, see Proposition 6.3 of [ČTW11]. An important fact about random interlacements on $\mathbb{T}^d$ is that the law of the vacant cluster of the root $\varnothing \in \mathbb{T}^d$ under Q^u is the same as the law of a certain (inhomogeneous) Galton-Watson tree. More precisely, the probability that $\varnothing$ is vacant equals $e^{-u\operatorname{cap}(\varnothing)} = \exp\{-u\frac{d-2}{d-1}\}$. Given that $\varnothing \in \mathcal{V}^u_{\mathbb{T}^d}$, the offspring distribution of the root is binomial with parameters d and

$$p_u = \exp\Big\{ -u\frac{(d-2)^2}{d(d-1)}\Big\}, \tag{5.2}$$

and the offspring distribution of all remaining individuals is binomial with parameters $d-1$ and p_u. Using this characterisation it is easy to compute the probability that the degree of the root in $\mathcal{V}^u_{\mathbb{T}^d}$ is i, $i = 0, \dots, d$.

$$\begin{aligned} d^u_i &:= Q^u[\text{degree of } \varnothing \text{ in } \mathcal{V}^u_{\mathbb{T}^d} \text{ equals } i] \\ &= Q^u[\varnothing \text{ is vacant and has exactly } i \text{ offsprings}] \\ &= e^{-u\frac{d-2}{d-1}} \binom{d}{i} p_u^i (1-p_u)^{d-i}. \end{aligned} \tag{5.3}$$

We now explain the relation between d^u_i and the degree sequence of $\mathcal{V}^u_G$. This relation was already obtained in a weaker form in Theorem 3 of [CF10] and could also be extracted from [ČTW11] Proposition 6.3. However the finer control of errors obtained in Theorem 5.1 is crucial if one wants to stay inside the critical window.

Recall from Section 3.1 that $\mathcal{D}^u$ denotes the degree sequence of the vacant graph $\mathbf{V}^u$ and that, for any degree sequence $\boldsymbol{d}$, $n_i(\boldsymbol{d})$ denotes the number of vertices with degree i in $\boldsymbol{d}$.

Theorem 5.1. *For every $u > 0$ and every $i \in \{0, \dots, d\}$,*

$$\big|E^G[n_i(\mathcal{D}^u)] - nd^u_i\big| \leq c(\log^5 n)n^{1/2}, \quad \bar{\mathbb{P}}_{n,d}\text{-a.a.s.} \tag{5.4}$$

Proof. We fix $x \in V_n$, $D \subset [d]$, and recall from Section 4 the definitions of the covering map ϕ, of $B_{x,D}$, $\mathbb{B}_D$ and of z_i. To simplify the notation we use σ and B as shorthand for $\sigma_{B_{x,D}}$ and $B_{x,D}$. We first estimate the probability that $B \subset \mathcal{V}^u$. Using the Markov property and Lemma 4.4, we obtain that $\mathbb{P}_{n,d}$-a.a.s.

$$\begin{aligned} &\Big| P^G[H_B > un] - \exp\Big\{ -\frac{un}{E^G_\sigma[H_B]}\Big\}\Big| \\ (5.5) \qquad &= \Big| P^G[H_B > t_n] E^G\big[P^G_{X_{t_n}}[H_B > un - t_n]\big| H_B > t_n\big] - \exp\Big\{-\frac{un}{E^G_\sigma[H_B]}\Big\}\Big| \\ &\le \Big| P^G[H_B > t_n] P^G_\sigma[H_B > un - t_n] - \exp\Big\{-\frac{un}{E^G_\sigma[H_B]}\Big\}\Big| + ce^{-c'\log^2 n}. \end{aligned}$$

Under the measure P_σ, the random variable H_B is exponentially distributed, see for instance [AB93] below (12). Hence, using that e^{-t} is 1-Lipschitz for $t \ge 0$,

$$(5.6) \qquad \Big| P^G_\sigma[H_B > un - t_n] - \exp\Big\{-\frac{un}{E_\sigma[H_B]}\Big\}\Big| \le \frac{t_n}{E^G_\sigma[H_B]}.$$

Therefore, (5.5) becomes

$$(5.7) \qquad \Big| P^G[H_B > un] - \exp\Big\{-\frac{un}{E^G_\sigma[H_B]}\Big\}\Big| \le ce^{-c'\log^2 n} + \frac{t_n}{E_\sigma[H_B]} + P[H_B \le t_n].$$

Let $V_{\text{good}} \subset V$ be the set of vertices $x \in V$ satisfying

$$(5.8) \qquad \begin{gathered} B(x,2) \text{ is a tree, and for every } D \subset [d] \text{ and } i \in D, \\ \big| P^G_{\phi(z_i)}[\tilde H_{B_{x,D}} > t_n] - e_{\mathbb{B}}(\phi(z_i))\big| \le n^{-1/2} \text{ and } P^G[H_{B_{x,D}} \le t_n] \le n^{-1/2}. \end{gathered}$$

By Corollary 4.3 above, and by Remark 1.4 and Lemma 6.1 of [ČTW11], the complement of V_{good} is very small,

$$(5.9) \qquad |V \setminus V_{\text{good}}| \le cn^{1/2}(\log n)^5, \qquad \mathbb{P}_{n,d}\text{-a.a.s.}$$

By Lemma 2 of [AB93] and (4.21)

$$(5.10) \qquad E^G_\sigma[H_B]^{-1} \le E^G[H_B]^{-1} \le cn^{-1}.$$

Therefore for $x \in V_{\text{good}}$, (5.7) becomes

$$(5.11) \qquad \Big| P^G[H_B > un] - \exp\Big\{-\frac{un}{E^G_\sigma[H_B]}\Big\}\Big| \le cn^{-1/2}.$$

Our next step is to obtain an estimate on $E_\sigma[H_B]$ for $x \in V_{\text{good}}$. To this aim we 'collapse' the set $B_{x,D}$ into one point b, and define a new Markov chain whose distribution is denoted $\bar P$ and which is characterised by its transition rates $\bar p_{xy}$,

$$(5.12) \qquad \begin{cases} \bar p_{ww'} = p_{ww'}, & \text{if } w, w' \ne b, \\ \bar p_{wb} = \sum_{y\in B} p_{wy}, & \text{if } w \ne b, \\ \bar p_{bw} = \frac{1}{|B|}\sum_{y\in B} p_{yw}, & \text{if } w \ne b. \end{cases}$$

It is easy to check that $\bar\pi = \frac{1}{n}\big(|B|\delta_b + \sum_{x\ne b}\delta_x\big)$ is a reversible distribution for this chain. Therefore,

$$(5.13) \qquad \frac{n}{|B|} = \bar E_b[\tilde H_b] = \bar E_b[\tilde H_b, \tilde H_b \le t_n] + \bar P_b[\tilde H_b > t_n]\bar E_b\big[\bar E_{X_{t_n}}[H_b - t_n]\big|\tilde H_b > t_n\big].$$

By (5.12), $\bar{P}_b[\tilde{H}_b > t_n] = |B|^{-1} \sum_{y\in B} P_y^G[\tilde{H}_B > t_n]$. Therefore, using Lemma 4.4 and (5.13),

$$\Big|\sum_{y\in B} P_y^G[\tilde{H}_B > t_n] E_\sigma^G[H_B] - n\Big| \le ct_n + c\exp\{-c' t_n\} \le ct_n. \tag{5.14}$$

Using (5.10), this yields the following estimate on $E_\sigma^G[H_B]$,

$$\Big|\sum_{y\in B} P_y^G[\tilde{H}_B > t_n] - \frac{n}{E_\sigma^G[H_B]}\Big| \le \frac{ct_n}{E_\sigma^G[H_B]} + c\exp\{-c' t_n\} \le \frac{c\log^2 n}{n}. \tag{5.15}$$

We are now in position to give our final estimate on $P^G[H_B \ge un]$. By the triangle inequality, for $x \in V_{\text{good}}$,

$$\begin{aligned}
&\Big|P^G[H_B > un] - \exp\Big\{-u\sum_{y\in\mathbb{B}} e_{\mathbb{B}}(y)\Big\}\Big| \\
&\quad\le \Big|P^G[H_B > un] - \exp\Big\{-\frac{un}{E_\sigma[H_B]}\Big\}\Big| \\
&\qquad + \Big|\exp\Big\{-\frac{un}{E_\sigma[H_x]}\Big\} - \exp\Big\{-u\sum_{y\in B} P_y^G[\tilde{H}_{B_{x,D}} > t_n]\Big\}\Big| \\
&\qquad + \Big|\exp\Big\{-u\sum_{y\in B_{x,D}} P_y^G[\tilde{H}_{B_{x,D}} > t_n]\Big\} - \exp\Big\{-u\sum_{y\in\mathbb{B}} e_{\mathbb{B}}(y)\Big\}\Big| \\
&\quad\le cn^{-1/2},
\end{aligned} \tag{5.16}$$

where for the last inequality we used the estimates (5.11), (5.15) and (5.8).

If $x \in V_{\text{good}}$, all its neighbours are distinct. We can then use the inclusion-exclusion formula to write

$$\begin{aligned}
P^G\big[\mathcal{D}^u(x) = i\big] &= \sum_{C\subset[d],|C|=i} P^G[\mathcal{V}^u \cap B(x,1) = B_{x,C}] \\
&= \sum_{C\subset[d],|C|=i}\ \sum_{D\subset[d],C\subset D} (-1)^{|D|-|C|} P^G\big[H_{B_{x,D}} > un\big].
\end{aligned} \tag{5.17}$$

Using (5.16), we obtain that

$$\Big|P^G\big[\mathcal{D}^u(x) = i\big] - \sum_{C\subset[d],|C|=i}\ \sum_{D\subset[d],C\subset D} (-1)^{|D|-|C|} \exp\Big\{-u\sum_{y\in\mathbb{B}_D} e_{\mathbb{B}_D}(y)\Big\}\Big| \le cn^{-1/2}. \tag{5.18}$$

From (4.1),(4.2), it is not difficult to deduce that for $y \in B_{x,D} \setminus \{x\}$,

$$e_{\mathbb{B}_D}(y) = \frac{d-1}{d}\cdot\frac{d-2}{d-1} \qquad \text{and} \qquad e_{\mathbb{B}_D}(x) = \frac{d-|D|}{d}\cdot\frac{d-2}{d-1}. \tag{5.19}$$

Inserting this into (5.18) leads to

(5.20)

$$\Big|P^G\big[\mathcal{D}^u(x) = i\big] - \binom{d}{i}\sum_{j=1}^{d}(-1)^{j-i}\binom{d-i}{j-i}\exp\Big\{-u\frac{d-2}{d-1}\Big(j\frac{d-1}{d} + \frac{d-j}{d}\Big)\Big\}\Big| \le cn^{-1/2}.$$

A simple computation implies that the leading term in the last formula equals d_i^u (see (5.3)). Therefore, $\bar{\mathbb{P}}_{n,d}$-a.a.s., uniformly for $x \in V_{\text{good}}$,

$$\big|P^G\big[\mathcal{D}^u(x) = i\big] - d_i^n\big| \le cn^{-1/2}. \tag{5.21}$$

The claim of Theorem 5.1 then follows by summing this relation over $x \in V_{\mathrm{good}}$ and using (5.9). □

We now prove the concentration of $n_i(\mathcal{D}^u)$ around its mean.

Theorem 5.2. *Let G be a d-regular (multi)graph on n vertices satisfying $\lambda_G \geq \alpha > 0$. Then, for every $\varepsilon \in (0, \frac{1}{4})$, and for every $i \in \{0, \dots, d\}$,*

$$P^G\big[|n_i(\mathcal{D}^u) - E^G[n_i(\mathcal{D}^u)]| \geq n^{1/2+\varepsilon}\big] \leq c_\varepsilon e^{-cn^\varepsilon}. \tag{5.22}$$

To prove this lemma we use the following concentration theorem for (not necessarily independent) random variables that we learnt from [McD98]. Consider a sequence $W = (W_1, \dots, W_M)$ of random variables, all taking values in some space $\mathcal{A}$. Let $f : \mathcal{A}^M \to \mathbb{R}$ be a bounded function. For $k \in \{1, \dots, M\}$ and $y_1, \dots, y_{k-1} \in \mathcal{A}^{k-1}$ we define

$$\begin{aligned} &r_k(y_1, \dots, y_{k-1}) \\ &\quad = \sup_{y,y' \in \mathcal{A}} \big|\mathbb{E}[f(W)|W_k = y, W_i = y_i \forall i < k] - \mathbb{E}[f(W)|W_k = y', W_i = y_i \forall i < k]\big| \end{aligned} \tag{5.23}$$

and set

$$R^2 = \sup\Big\{\sum_{k=1}^M r_k^2(y_1, \dots, y_{k-1}) : y_1, \dots, y_{M-1} \in \mathcal{A}\Big\}. \tag{5.24}$$

Lemma 5.3 (Theorem 3.7 of [McD98])**.** *Let $W = (W_1, \dots, W_M)$ be as above. Then*

$$\mathbb{P}[|f(W) - \mathbb{E}f(W)| \geq t] \leq 2e^{-2t^2/R^2}. \tag{5.25}$$

Proof of Theorem 5.2. To apply Lemma 5.3, we need the following construction similar to Section 4 of [ČTW11]. Let $\ell = n^\varepsilon$ for ε from the statement of Theorem 5.2. On an auxiliary probability space (Ω, Q), define $(Z^i, i \in \mathbb{N})$ to be a collection of i.i.d. uniformly chosen vertices of G. Given the collection (Z_i), let $(Y_i : i \geq 1)$ be (conditionally) independent family of elements of $D([0, \ell], G)$ such that Y^i is distributed according to the random walk bridge $P^{G,\ell}_{Z^{i-1},Z^i}$ (see Section 2.3 for the definition). We define $\mathcal{X} \in D([0, \infty), G)$ to be the concatenation of Y^i's,

$$\mathcal{X}(t) = Y^i(t - (i-1)\ell), \qquad \text{when } (i-1)\ell \leq t < i\ell. \tag{5.26}$$

We use $\mathcal{P}^G$ to denote the distribution of $\mathcal{X}$ on $D([0, \infty), G)$, $\mathcal{P}^G = Q \circ \mathcal{X}^{-1}$. $\mathcal{P}^{G,T}$ stands for its restriction to $D([0, T], G)$. The measure $\mathcal{P}^{G,un}$ approximates well $P^{G,un}$ if ℓ is large enough as follows from the next lemma whose proof is postponed to the end of this section.

Lemma 5.4. *Assume that $\lambda_G > \alpha$ and $\ell = n^\varepsilon$. Then there exist constant $c_{\alpha,\varepsilon}$ and $c'_{\alpha,\varepsilon}$ such that for every $u > 0$ and all n satisfying $ne^{-\ell\alpha} < 1/2$, $\mathcal{P}^{G,un}$ and $P^{G,un}$ are equivalent and*

$$\Big|\frac{\mathrm{d}\mathcal{P}^{G,nu}}{\mathrm{d}P^{G,nu}} - 1\Big| \leq c'_{\alpha,\epsilon} u e^{-c_{\alpha,\epsilon}\ell}. \tag{5.27}$$

To be able to apply Lemma 5.3, more precisely to estimate the functions r_k, we do not want $|\operatorname{Ran} Y^i|$ to be too large. Therefore, we define $\bar{Y}^i \subset V$ to be the set of first 2ℓ vertices visited by Y^i,

$$\bar{Y}^i = \{Y^i_t, t \leq (\tau_{2\ell}(Y^i) \wedge \ell)\}, \tag{5.28}$$

where $\tau_k(Y^i)$ denotes the time of k-th step Y^i (defined to be infinite if Y^i makes less than k steps). Obviously $\bar{Y}^i \subset \operatorname{Ran} Y^i$. On the other hand, it can be proved as in Lemma 4.2 of

[ČTW11], that

$$Q(\operatorname{Ran} Y^i \neq \bar{Y}^i) \leq \sup_{x,y\in V} P^{G,\ell}_{xy}[N_\ell \geq 2\ell] \leq ce^{-c'\ell}. \tag{5.29}$$

(Remark that Y^i has the law of the random walk bridge, and thus (5.29) is not just a direct consequence of large deviation estimate for a Poisson random variable.)

We may now prove Theorem 5.2. Set $m = \lfloor un/\ell \rfloor$ and $u' = m\ell/n$. Let N be the number of steps of X between $u'n$ and un. Since $un - u'n \leq \ell$, by properties of Poisson random variables, $P^G[N \geq 2\ell] \leq e^{-c\ell}$. Between, $u'n$ and un the walk visits at most N sites, therefore

$$|n_i(\mathcal{D}^u) - n_i(\mathcal{D}^{u'})| \leq (d+1)N, \tag{5.30}$$

and

$$|E^G[n_i(\mathcal{D}^u)] - E^G[n_i(\mathcal{D}^{u'})]| \leq (d+1)E^G[N] \leq (d+1)\ell. \tag{5.31}$$

Thus, for $n \geq c_{\alpha,\epsilon}$,

$$\begin{aligned} &P^G\big[|n_i(\mathcal{D}^u) - E^G[n_i(\mathcal{D}^u)]| \geq n^{1/2+\varepsilon}\big] \\ &\quad \leq P^G\big[|n_i(\mathcal{D}^{u'}) - E^G[n_i(\mathcal{D}^{u'})]| \geq n^{1/2+\varepsilon} - 3(d+1)\ell\big] + P^G[N \geq 2\ell]. \\ &\quad \leq P^G\big[|n_i(\mathcal{D}^{u'}) - E^G[n_i(\mathcal{D}^{u'})]| \geq \tfrac{1}{2}n^{1/2+\varepsilon}\big] + e^{-c\ell}. \end{aligned} \tag{5.32}$$

Using Lemma 5.4, denoting by $\mathcal{E}^G$ the expectation corresponding to $\mathcal{P}^G$,

$$\big|E^G[n_i(\mathcal{D}^{u'})] - \mathcal{E}^G[n_i(\mathcal{D}^{u'})]\big| \leq c_{\alpha,\epsilon} n u' e^{-c'_{\alpha,\epsilon}\ell}, \tag{5.33}$$

and thus, for $n \geq c_{\alpha,\epsilon}$,

$$\begin{aligned} &P^G\big[|n_i(\mathcal{D}^{u'}) - E^G[n_i(\mathcal{D}^{u'})]| \geq \tfrac{1}{2}n^{1/2+\varepsilon}\big] \\ &\quad \leq \mathcal{P}^G\big[|n_i(\mathcal{D}^{u'}) - \mathcal{E}^G[n_i(\mathcal{D}^{u'})]| \geq \tfrac{1}{4}n^{1/2+\varepsilon}\big] + c_{\alpha,\epsilon} u' e^{-c'_{\alpha,\epsilon}\ell}. \end{aligned} \tag{5.34}$$

Observe that under $\mathcal{P}^G$, $\mathcal{V}^{u'} = V \setminus \cup_{i\leq m} \operatorname{Ran} Y^i$. Let $\bar{\mathcal{V}}$ be the vacant set left by $\bar{Y}^i$'s, $\bar{\mathcal{V}} = V \setminus \cup_{i\leq m}\bar{Y}^i$, and denote by $\bar{\mathcal{D}}$ the degree sequence of the graph with set of vertices V and edge set $\{\{x,y\} \in \mathcal{E} : x,y \in \bar{\mathcal{V}}\}$, cf. (3.1). By (5.29), we have then

$$Q[\mathcal{D}^{u'} \neq \bar{\mathcal{D}}] \leq cme^{-c'\ell} \leq ce^{-c''\ell}. \tag{5.35}$$

Therefore, for $n \geq c_\alpha$,

$$\big|\mathcal{E}^G[n_i(\mathcal{D}^{u'})] - Q[n_i(\bar{\mathcal{D}})]\big| \leq nQ[\mathcal{D}^{u'} \neq \bar{\mathcal{D}}] \leq ce^{-c'\ell}, \tag{5.36}$$

and thus

$$\begin{aligned} &\mathcal{P}^G\big[|n_i(\mathcal{D}^{u'}) - \mathcal{E}^G[n_i(\mathcal{D}^{u'})]| \geq \tfrac{1}{4}n^{1/2+\varepsilon}\big] \\ &\quad \leq Q\big[|n_i(\bar{\mathcal{D}}) - Q[n_i(\bar{\mathcal{D}})]| \geq \tfrac{1}{8}n^{1/2+\varepsilon}\big] + ce^{-c'\ell}. \end{aligned} \tag{5.37}$$

We now apply Lemma 5.3 with $M = m$, $W_i = \bar{Y}^i$, $f = n_i(\bar{\mathcal{D}})$, and $\mathcal{A}$ being the set of subsets of V with at most 2ℓ elements. Writing $\boldsymbol{y}_k = (y_1,\dots,y_k)$, $\boldsymbol{y}'_k = (y_1,\dots,y_{k-1},y'_k)$, and $\boldsymbol{Y}_k = (\bar{Y}_1,\dots,\bar{Y}_k)$, we claim that

$$r_k(\boldsymbol{y}_{k-1}) = \sup_{y,y'\in\mathcal{A}} |Q[n_i(\bar{\mathcal{D}})|\boldsymbol{Y}_k = \boldsymbol{y}_k] - Q[n_i(\bar{\mathcal{D}})|\boldsymbol{Y}_k = \boldsymbol{y}'_k]| \leq 2(d+1)\ell. \tag{5.38}$$

Indeed, by conditioning also on the values of $\bar{Y}^{k+2},\dots,\bar{Y}^m$, we observe that the difference

$$|Q[n_i(\bar{\mathcal{D}})|\boldsymbol{Y}_k = \boldsymbol{y}_k, \bar{Y}^{k+2},\dots,\bar{Y}^m] - Q[n_i(\bar{\mathcal{D}})|\boldsymbol{Y}_k = \boldsymbol{y}'_k, \bar{Y}^{k+2},\dots,\bar{Y}^m]| \tag{5.39}$$

cannot be larger than $(d+1)(|\bar{Y}^k| + |\bar{Y}^{k+1}|) \leq 2(d+1)\ell$. The inequality (5.38) then follows by integrating over $\bar{Y}^{k+2},\dots,\bar{Y}^m$.

From (5.38) it follows that we can apply Lemma 5.3 with $R^2 = m(d+1)^2\ell^2 = cn^{1+\varepsilon}$, yielding

$$Q\big[|n_i(\bar{\mathcal{D}}) - Q[n_i(\bar{\mathcal{D}})]| \ge \tfrac{1}{8} n^{1/2+\varepsilon}\big] \le ce^{-cn^{1+2\varepsilon}/n^{1+\varepsilon}} \le ce^{-cn^{\varepsilon}}. \tag{5.40}$$

This, together with (5.32), (5.34) and (5.37) completes the proof of Theorem 5.2. □

Proof of Lemma 5.4. Let u' be the smallest number greater or equal to u, such that $u'n$ is an integer multiple of ℓ, and set $m = u'n/\ell$. Let further A be an arbitrary $\mathcal{F}_{un}$-measurable subset of $D([0,u'n],V)$. Since $P^{G,un}$ and $\mathcal{P}^{G,un}$ are the restrictions of the measures $P^{G,u'n}$ and $\mathcal{P}^{G,u'n}$ to $D([0,un],V)$, it is sufficient to prove the lemma with u replaced by u'. To this end we write

$$P^{G,u'n}[A] = \sum_{x_0,\dots,x_m \in V} P^{G,u'n}[A | X_{i\ell} = x_i, 0 \le i \le m] P^{G,u'n}[X_{i\ell} = x_i, 0 \le i \le m]. \tag{5.41}$$

By the Markov property

$$P^{G,u'n}[X_{i\ell} = x_i, 0 \le i \le m] = \pi(x_0) \prod_{k=0}^{m-1} P^{\ell}_{x_k}[X_\ell = x_{k+1}]. \tag{5.42}$$

The construction of the measure $\mathcal{P}^{G,u'n}$ implies that

$$\begin{aligned} \mathcal{P}^{G,u'n}[A | X_{i\ell} = x_i, 0 \le i \le m] &= P^{G,u'n}[A | X_{i\ell} = x_i, 0 \le i \le m], \\ \mathcal{P}^{G,u'n}[X_{i\ell} = x_i, 0 \le i \le 2m] &= \prod_{k=0}^{m} \pi(x_k). \end{aligned} \tag{5.43}$$

Comparing (5.42) and (5.43), it remains to control the ratio $P^{\ell}_x[X_\ell = y]/\pi(y)$. However, by (2.9) and the assumption of the lemma, $|P^{G,\ell}_x[X_\ell = y]/\pi(y) - 1| \le ne^{-\alpha\ell}$. This leads to

$$(1 - ne^{\alpha\ell})^m \le \frac{\mathcal{P}^{G,u'n}[A]}{P^{G,u'n}[A]} \le (1 + ne^{-\alpha\ell})^m \tag{5.44}$$

Since $\ell = n^{\varepsilon}$ and $ne^{-\ell\alpha} \le \frac{1}{2}$ by the assumptions of the lemma, it immediately follows that $P^{G,u'n}$ and $\mathcal{P}^{G,u'n}$ are equivalent. A change of constants accommodating the terms polynomial in n then completes the proof. □

6. Proofs of Theorems 1.1 and 1.2

We now have all tools that we need to show all the main results of this paper. As a direct consequence of Theorems 5.1, 5.2 and the fact (2.10), we get that $\bar{\mathbf{P}}_{n,d}$-a.a.s.

$$|n_i(\mathcal{D}^u) - nd^u_i| \le cn^{1/2} \log^5 n, \qquad \text{for all } 0 \le i \le d, \tag{6.1}$$

where d^u_i is defined in (5.3). Hence, $\bar{\mathbf{P}}_{n,d}$-a.a.s, $\lim_{n\to\infty} n^{-1} n_i(\mathcal{D}^u) = d^u_i$. The constant $Q(\mathcal{D}^u)$ (see (3.6)) can be written as

$$Q(\mathcal{D}^u) = \frac{\sum_{x=1}^n \mathcal{D}^u(x)^2}{\sum_{x=1}^n \mathcal{D}^u(x)} - 2 = \frac{\sum_{i=1}^d i^2 n_i(\mathcal{D}^u)}{\sum_{i=1}^d i n_i(\mathcal{D}^u)} - 2. \tag{6.2}$$

Thus, for $n > c_u$, on the event in (6.1) we have

$$\Big| Q(\mathcal{D}^u) - \big(p_u(d-1) - 1\big) \Big| \le c_u n^{-1/2} \log^5 n. \tag{6.3}$$

The value $u_\star$ given by (1.2) satisfies $p_{u_\star}(d-1)-1=0$. Therefore, when $u_n \to u_\star$, we obtain by expanding the exponential in the definition (5.2) of p_u around $u_\star$,

$$\Big|Q(\mathcal{D}^{u_n}) - (u_\star - u_n)\frac{(d-2)^2}{d(d-1)}\Big| \le c\big((u_\star - u_n)^2 + n^{-1/2}\log^5 n\big). \tag{6.4}$$

This implies that when u_n is in the critical window of Theorem 1.1, that is $|n^{1/3}(u_\star - u_n)| \le \lambda$, then $Q(\mathcal{D}^{u_n})$ is in the critical window of Theorem 3.2, that is $n^{1/3}|Q(\mathcal{D}^{u_n})| \le \lambda'$, $\bar{\mathbf{P}}_{n,d}$-a.a.s. Theorem 1.1 then follows directly from Theorem 3.2(i) together with Proposition 3.1 and the remark following (2.3).

Very similar reasoning apply when proving Theorems 1.2 and 1.3. We should only identify the constants of Theorem 3.2. Easy computations give

$$\lambda = e^{-u\frac{d-2}{d-1}} dp_u, \qquad \beta = e^{-u\frac{d-2}{d-1}} d(d-1)(d-2)p_u^3, \tag{6.5}$$

and thus

$$v_n = 2n\lambda^2\beta^{-1}Q(\mathcal{D}^{u_n}) = 2n(u_\star - u_n)\frac{d-2}{(d-1)^2}e^{-u_\star\frac{d-2}{d-1}}(1+o(1)). \tag{6.6}$$

Replacing $u_\star - u_n$ by $\omega_n n^{-1/3}$, Theorem 1.2(a) follows. It can also be seen that $\sqrt{n/Q(\mathcal{D}^{u_n})}$ is of order $n^{2/3}\omega_n^{-1/2}$, implying Theorem 1.2(b).

Finally to identify ρ of Theorem 1.3. We observe that $g(x)$ of Theorem 3.2 is given by

$$g(x) = \sum_{i=0}^{d} d_i^u x^i = e^{-u\frac{d-2}{d-1}}(xp_u + (1-p_u))^d. \tag{6.7}$$

After few simplifications, ξ of Theorem 3.2 is the unique solution in $(0,1)$ of the equation

$$(xp_u + (1-p_u))^{d-1} = x, \tag{6.8}$$

and ρ is given by

$$\rho = 1 - g(\xi). \tag{6.9}$$

This completes the proofs of all three main theorems.

Remark 6.1. (1) Theorem 5.1 raises the question of what is the right magnitude of deviations in $E^G[n_i(\mathcal{D}^u)]$ under the law $\bar{\mathbb{P}}_{n,d}$? If indeed it is of order $n^{1/2}$ (without power-log corrections), then it would be interesting to investigate whether this quantity satisfies a central limit theorem when properly rescaled.

(2) As established in Proposition 3.1 and Theorems 5.1 and 5.2, we can reduce the study of $\mathcal{V}^u$ to questions on the behaviour of random graphs with prescribed degree sequences. Although the results in [MR95, JL09, HM10] provide very fine information about such graphs, there are several questions concerning them which are still open. For instance, one could give a better description of the geometry of their critical components, their diameters, spectral gaps, etc.

(3) It is interesting to notice that the statements (1.3) and (1.4) were established in [ČTW11] for the vacant set left by random walk on other sequences of graphs, such as large girth expanders. Is it possible to extend the results of the current paper on the critical behaviour of $\mathcal{V}^u$ to this more general setting?

Appendix A. Properties of the quasi-stationary distribution

We establish here few results for arbitrary reversible irreducible continuous-time Markov chains on a finite state space. These results are natural but we have not found any suitable formulation in the literature.

Let V be a finite set and let $\mathcal{L}$ be the generator of a reversible irreducible continuous-time Markov chain X on V, and let $\pi(x)$ be its invariant measure. We use $\langle f, g\rangle$ to denote the usual scalar product on $L^2(V,\pi)$, $\langle f, g\rangle = \sum_{x\in V} f(x)g(x)\pi(x)$. The operator $-\mathcal{L}$ is symmetric in $L^2(V,\pi)$ and has real eigenvalues $0 = \lambda_1 < \lambda_2 \le \cdots \le \lambda_{|V|}$ and corresponding orthonormal eigenvectors $v_1, \dots, v_{|V|}$.

For $B \subset V$, we use $\mathcal{L}^B$ to denote the generator of the Markov chain X killed on hitting B, viewed as an operator on $L^2(V\setminus B, \pi|_{V\setminus B})$. Let $0 < \lambda_1^B < \lambda_2^B \le \dots\dots \lambda^B_{|V\setminus B|}$, and v_1^B, $\dots$, $v^B_{|V\setminus B|}$ denote the eigenvalues and eigenvectors of $-\mathcal{L}^B$.

The *quasi-stationary distribution* σ_B is related to the eigenvector of $-\mathcal{L}^B$ corresponding to λ_1^B and is given by

$$\sigma_B(y) = \frac{v_1^B(y)\pi(y)}{\langle v_1^B, \mathbf{1}\rangle} = \frac{\langle v_1^B, \delta_y\rangle}{\langle v_1^B, \mathbf{1}\rangle}, \tag{A.1}$$

with $\mathbf{1}$ denoting the constant one function. Inverting this relation we get

$$v_1^B(x) = \frac{\sigma_B(x)}{\pi(x)}\Big(\sum_{x\in V\setminus B}\frac{\sigma_B(x)^2}{\pi(x)}\Big)^{-1/2}. \tag{A.2}$$

Lemma A.1. *For every $B\subset V$,*

$$\lambda_2^B - \lambda_1^B \ge \lambda_2 - \frac{1}{E[H_B]}. \tag{A.3}$$

Proof of Lemma A.1. Since $\mathcal{L}^B$ can be viewed as a sub-matrix of $\mathcal{L}$, by the eigenvalue interlacing inequality (cf. [Hae95], Corollary 2.2), we have $\lambda_2^B \ge \lambda_2$. On the other hand, by [AB93] Lemma 2 and the paragraph following equation (12),

$$\lambda_1^B = \frac{1}{E_{\sigma_B}[H_B]} \le \frac{1}{E[H_B]}. \tag{A.4}$$

Combining these two inequalities we obtain Lemma A.1. □

Lemma A.2. *Suppose that for $t > 0$ and $\varepsilon \in (0, 1/2)$*

$$e^{-t(\lambda_2^B-\lambda_1^B)}|V\setminus B|\Big(\sup_{x\in V\setminus B}\frac{\sigma_B(x)}{\sqrt{\pi(x)}}\Big)^2 \le \varepsilon \inf_{x\in V\setminus B}\frac{\sigma_B(x)}{\sqrt{\pi(x)}}. \tag{A.5}$$

Then,

$$\sup_{x,y\in V\setminus B}\big|P_x[X_t = y | H_B > t] - \sigma_B(y)\big| \le 4\varepsilon. \tag{A.6}$$

Proof. In the proof we will only use the eigenvalues and eigenvectors of $-\mathcal{L}^B$, therefore we omit the superscript B from the notation. Similarly, we write σ for σ_B and define $m = |V\setminus B|$. By the usual spectral decomposition formula,

$$\begin{aligned} P_x[X_t = y, H_B > t] &= (e^{t\mathcal{L}^B}\delta_y)(x) = \sum_{k=1}^m e^{-\lambda_k t} v_k(x)\langle v_k, \delta_y\rangle,\\ P_x[H_B > t] &= (e^{t\mathcal{L}^B}\mathbf{1})(x) = \sum_{k=1}^m e^{-\lambda_k t} v_k(x)\langle v_k, \mathbf{1}\rangle, \end{aligned} \tag{A.7}$$

where δ_y is the indicator function of y. For $f \in L^2(\pi)$, define $\psi_f = \sum_{k=2}^m e^{-(\lambda_k-\lambda_1)t}\langle v_k, f\rangle v_k$. Then $e^{t\mathcal{L}^B} f = e^{-\lambda_1 t}(v_1\langle v_1, f\rangle + \psi_f)$, and by Pythagoras' theorem

$$\|\psi_f\|_{L^2(\pi)} \le e^{-(\lambda_2-\lambda_1)t}\|f\|_{L^2(\pi)}. \tag{A.8}$$

Using this notation and the definition of the conditional probability,

$$P_x[X_t = y | H_B > t] = \frac{v_1(x)\langle v_1, \delta_y\rangle + \psi_{\delta_y}(x)}{v_1(x)\langle v_1, \mathbf{1}\rangle + \psi_{\mathbf{1}}(x)}. \tag{A.9}$$

Applying (A.1) we get after an easy algebra

$$P_x[X_t = y | H_B > t] - \sigma(y) = \frac{\psi_{\delta_y}(x) - \frac{\langle v_1, \delta_y\rangle}{\langle v_1, \mathbf{1}\rangle}\psi_{\mathbf{1}}(x)}{v_1(x)\langle v_1, \mathbf{1}\rangle(1 + \frac{\psi_{\mathbf{1}}(x)}{v_1(x)\langle v_1, \mathbf{1}\rangle})}. \tag{A.10}$$

Let f stand either for δ_y or $\mathbf{1}$. Then $\|f\|_{L^2(\pi)} \le 1$, which directly implies $|\psi_f(x)| \le \pi(x)^{-1/2}e^{-(\lambda_2-\lambda_1)t}$. From (A.8), (A.2), using the assumption (A.5), we obtain

$$\frac{\psi_f(x)}{v_1(x)\langle v_1, \mathbf{1}\rangle} \le \frac{e^{-(\lambda_2-\lambda_1)t}\sum_z \frac{\sigma(z)^2}{\pi(z)}}{\frac{\sigma(x)}{\sqrt{\pi(x)}}} \le \frac{e^{-(\lambda_2-\lambda_1)t} m \sup_z \left(\frac{\sigma(z)}{\sqrt{\pi(z)}}\right)^2}{\inf_z \frac{\sigma(z)}{\sqrt{\pi(z)}}} \le \varepsilon. \tag{A.11}$$

Using $\varepsilon < 1/2$, this implies that the absolute value of (A.10) can be bounded from above by

$$\frac{2(|\psi_{\delta_y}(x)| + |\psi_{\mathbf{1}}(x)|)}{v_1(x)\langle v_1, \mathbf{1}\rangle} \le 4\varepsilon. \tag{A.12}$$

This completes the proof of Lemma A.2. □

References

[AB93] David J. Aldous and Mark Brown, *Inequalities for rare events in time-reversible Markov chains. II*, Stochastic Process. Appl. **44** (1993), no. 1, 15–25. MR 1198660

[ABS04] Noga Alon, Itai Benjamini, and Alan Stacey, *Percolation on finite graphs and isoperimetric inequalities*, Ann. Probab. **32** (2004), no. 3A, 1727–1745. MR 2073175

[BS08] Itai Benjamini and Alain-Sol Sznitman, *Giant component and vacant set for random walk on a discrete torus*, J. Eur. Math. Soc. (JEMS) **10** (2008), no. 1, 133–172. MR 2349899

[BF82] B. Bollobás and W. Fernandez de la Vega, *The diameter of random regular graphs*, Combinatorica **2** (1982), no. 2, 125–134. MR 685038

[Bol01] Béla Bollobás, *Random graphs*, second ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001. MR 1864966

[ČTW11] Jiří Černý, Augusto Teixeira, and David Windisch, *Giant vacant component left by a random walk in a random d-regular graph*, To appear in Ann. Inst. H. Poincaré Probab. Statist., 2011.

[CF10] Colin Cooper and Alan Frieze, *Component structure induced by a random walk on a random graph*, arXiv 1005.1564, 2010.

[FR09] Nikolaos Fountoulakis and Bruce Reed, *A general critical condition for the emergence of a giant component in random graphs with given degrees*, European Conference on Combinatorics, Graph Theory and Applications (EuroComb 2009), Electron. Notes Discrete Math., vol. 34, Elsevier Sci. B. V., Amsterdam, 2009, pp. 639–645. MR 2720906

[Fri08] Joel Friedman, *A proof of Alon's second eigenvalue conjecture and related problems*, Mem. Amer. Math. Soc. **195** (2008), no. 910, viii+100. MR 2437174

[Hae95] Willem H. Haemers, *Interlacing eigenvalues and graphs*, Linear Algebra Appl. **226/228** (1995), 593–616. MR 1344588

[HM10] Hamed Hatami and Michael Molloy, *The scaling window for a random graph with a given degree sequence*, Proceedings of SODA 2010, 2010, more exhaustive version of the paper appears at arXiv:0907.4211.

[JL09] Svante Janson and Malwina J. Luczak, *A new approach to the giant component problem*, Random Structures Algorithms **34** (2009), no. 2, 197–216. MR 2490288

[McD98] Colin McDiarmid, *Concentration*, Probabilistic methods for algorithmic discrete mathematics, Algorithms Combin., vol. 16, Springer, Berlin, 1998, pp. 195–248. MR 1678578

[MR95] Michael Molloy and Bruce Reed, *A critical point for random graphs with a given degree sequence*, Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science, "Random Graphs '93" (Poznań, 1993), vol. 6, 1995, pp. 161–179. MR 1370952

[NP10] Asaf Nachmias and Yuval Peres, *Critical percolation on random regular graphs*, 2010, pp. 111–148. MR 2583058

[Pit08] Boris Pittel, *Edge percolation on a random regular graph of low degree*, Ann. Probab. **36** (2008), no. 4, 1359–1389. MR 2435852

[SC97] Laurent Saloff-Coste, *Lectures on finite Markov chains*, Lectures on probability theory and statistics (Saint-Flour, 1996), Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, pp. 301–413. MR 1490046

[Szn10] Alain-Sol Sznitman, *Vacant set of random interlacements and percolation*, Ann. of Math. (2) **171** (2010), no. 3, 2039–2087. MR 2680403

[Tei09] A. Teixeira, *Interlacement percolation on transient weighted graphs*, Electron. J. Probab. **14** (2009), no. 54, 1604–1628. MR 2525105

[TW10] Augusto Teixeira and David Windisch, *On the fragmentation of a torus by random walk*, arXiv 1007.0902, 2010.

[Woe00] Wolfgang Woess, *Random walks on infinite graphs and groups*, Cambridge Tracts in Mathematics, vol. 138, Cambridge University Press, Cambridge, 2000. MR 1743100

[Wor99] N. C. Wormald, *Models of random regular graphs*, Surveys in combinatorics, 1999 (Canterbury), London Math. Soc. Lecture Note Ser., vol. 267, Cambridge Univ. Press, Cambridge, 1999, pp. 239–298. MR 1725006