

Level-set percolation of the Gaussian free field on regular graphs III: giant component on expanders

Jiří Černý

Departement Mathematik und Informatik
Fachbereich Mathematik
Universität Basel
CH-4051 Basel

Preprint No. 2021-14
June 2021

dmi.unibas.ch

LEVEL-SET PERCOLATION OF THE GAUSSIAN FREE FIELD ON REGULAR GRAPHS III: GIANT COMPONENT ON EXPANDERS

JIRÍ ČERNÝ

ABSTRACT. We consider the zero-average Gaussian free field on a certain class of finite d -regular graphs for fixed $d \geq 3$. This class includes d -regular expanders of large girth and typical realisations of random d -regular graphs. We show that the level set of the zero-average Gaussian free field above level h has a giant component in the whole supercritical phase, that is for all $h < h_*$, with probability tending to one as the size of the graphs tends to infinity. In addition, we show that this component is unique. This significantly improves the result of [AČ20b], where it was shown that a linear fraction of vertices is in mesoscopic components if $h < h_*$.

1. INTRODUCTION

In this paper we continue investigations of the level-set percolation of the zero-average Gaussian free field on a certain class of finite d -regular random graphs. These investigations were initiated in [AČ20b], after preparatory steps conducted in [AČ20a]. It was shown there that the model exhibits a percolation phase transition at a critical level h_* in the following sense: With probability tending to one as the size N of the graphs tends to infinity, whenever $h > h_*$, the level set of the zero-average Gaussian free field above level h does not contain any connected component of size larger than $C_h \log N$, and, on the contrary, whenever $h < h_*$, a linear fraction of the vertices is contained in ‘mesoscopic’ connected components of the level set above level h , that is in components having a size of at least a small fractional power of N . The critical level h_* agrees with the percolation threshold of the level set percolation of the usual Gaussian free field on a d -regular tree which was identified in [Szn16].

In the subcritical phase, $h > h_*$, this description of the behaviour of the level set is satisfactory. On the other hand, in the supercritical phase, $h < h_*$, it leaves open the question whether the mesoscopic components form a giant component, that is a component of size of order N , cf. [AČ20b, Remark 5.7].

This is a natural question since for other probabilistic models on essentially the same class of graphs the emergence of the unique giant component has been shown in the corresponding supercritical phases. The first example is the Bernoulli bond percolation on d -regular expanders of large girth considered in [ABS04] (for more recent results, see [KLS20]). A second example is the percolation of the vacant set left by the simple random walk on the same class of graphs as considered here and in [AČ20b], see [ČTW11]. In particular the latter result gives a strong indication that a giant component should emerge also in the supercritical phase of the level-set percolation, as the two models share many common features, like similar decay of correlations.

We answer this question affirmatively. To state our results we first recall the setting of [AČ20b]. We fix $d \geq 3$ and assume that $(\mathcal{G}_n)_{n \geq 1}$ is a sequence of graphs satisfying the following conditions.

Assumption 1.1. There exist $\alpha \in (0, 1)$, $\beta > 0$, and an increasing sequence of positive integers $(N_n)_{n \geq 1}$ with $\lim_{n \rightarrow \infty} N_n = \infty$ such that for all $n \geq 1$:

- (a) $\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$ is a simple connected graph with N_n vertices which is d -regular (that is all its vertices have degree d).
- (b) For all $x \in \mathcal{V}_n$ there is at most one cycle in the ball of radius $\lfloor \alpha \log_{d-1}(N_n) \rfloor$ around x .
- (c) The spectral gap of \mathcal{G}_n , denoted by $\lambda_{\mathcal{G}_n}$, satisfies $\lambda_{\mathcal{G}_n} \geq \beta$.

We refer to [AČ20b] for a more detailed discussion of these assumptions, but recall that they are satisfied for two important classes of graphs: (a) random d -regular graphs, (b) d -regular expanders of large girth.

On \mathcal{G}_n we consider the zero-average Gaussian free field $\Psi_{\mathcal{G}_n} = (\Psi_{\mathcal{G}_n}(x) : x \in \mathcal{V}_n)$ which is a centred Gaussian process on \mathcal{V}_n whose law is determined by its covariance function

$$E[\Psi_{\mathcal{G}_n}(x)\Psi_{\mathcal{G}_n}(y)] = G_{\mathcal{G}_n}(x, y) \quad \text{for all } x, y \in \mathcal{V}_n, \quad (1.1)$$

where $G_{\mathcal{G}_n}(\cdot, \cdot)$ is the zero-average Green function on \mathcal{G}_n (see (2.3), (2.5) for its definition).

The zero-average Gaussian free field is a natural version of the Gaussian free field for finite graphs. However, due to the zero-average property, namely

$$\sum_{x \in \mathcal{V}_n} \Psi_{\mathcal{G}_n}(x) = 0, \quad \text{a.s.}, \quad (1.2)$$

it comes with some peculiarities like the lack of an FKG-inequality and of the domain Markov property, cf. [AČ20b, Section 2.2].

We analyse the properties of the level sets of $\Psi_{\mathcal{G}_n}$ above level $h \in \mathbb{R}$, that is of

$$E^{\geq h}(\Psi_{\mathcal{G}_n}) := \{x \in \mathcal{V}_n : \Psi_{\mathcal{G}_n}(x) \geq h\}. \quad (1.3)$$

In particular, we are interested in the sizes of its largest and second largest connected components $\mathcal{C}_{\max}^{\mathcal{G}_n, h}$ and $\mathcal{C}_{\text{sec}}^{\mathcal{G}_n, h}$.

For our investigations it is important that the field $\Psi_{\mathcal{G}_n}$ is locally well approximated by the Gaussian free field $\varphi_{\mathbb{T}_d} = (\varphi_{\mathbb{T}_d}(x) : x \in \mathbb{T}_d)$ on the infinite rooted d -regular tree \mathbb{T}_d (see the paragraph containing (2.6) for the definition). For now, we only define its percolation function

$$\eta(h) := P(|\mathcal{C}_o^h| = \infty), \quad (1.4)$$

where \mathcal{C}_o^h is the connected component of the set $E^{\geq h}(\varphi_{\mathbb{T}_d}) := \{x \in \mathbb{T}_d : \varphi_{\mathbb{T}_d}(x) \geq h\}$ containing the root $o \in \mathbb{T}_d$, and we set

$$h_\star := \inf \{h \in \mathbb{R} : \eta(h) = 0\}, \quad (1.5)$$

to be its critical value. From [Szn16] it is known that h_\star is positive and finite.

We can now state our main result.

Theorem 1.2. *If $h < h_\star$, then for every sequence of graphs $(\mathcal{G}_n)_{n \geq 1}$ satisfying Assumption 1.1 and every $\delta > 0$*

$$\lim_{n \rightarrow \infty} P\left(\frac{|\mathcal{C}_{\max}^{\mathcal{G}_n, h}|}{N_n} \in (\eta(h) - \delta, \eta(h) + \delta) \text{ and } |\mathcal{C}_{\text{sec}}^{\mathcal{G}_n, h}| \leq \delta N_n\right) = 1. \quad (1.6)$$

Theorem 1.2 confirms the emergence of the giant component in the supercritical phase of the model, gives its typical size, and provides its uniqueness.

Similarly as in [ABS04, ČTW11, KLS20], we use a sprinkling technique to show that the mesoscopic components (that we know to exist due to [AČ20b, Theorem 5.1]) indeed form a giant component. Making the sprinkling work in the settings of dependent percolation is however rather challenging, as was already observed in [ČTW11], in the context of the vacant set left by a random walk. In the context of Gaussian free field, sprinkling was previously used in [DR15], to show the existence of an infinite connected component of the supercritical level set on \mathbb{Z}^d when $d \rightarrow \infty$. The diverging dimension is important for the arguments therein, since the correlations of the field decay with the dimension (as d^{-1} for the neighbouring vertices). Several sprinkling steps are also used

in the recent groundbreaking paper [DGRS20], which proves the sharpness of the phase transition for the level set of the Gaussian free field on \mathbb{Z}^d , $d \geq 3$. Note also that the results of [DGRS20] can be combined with [Abä19] to show the existence of the giant component for the supercritical level set of the zero-average Gaussian free field on a large discrete torus (cf. [DGRS20, Section 1.2]).

Very recently, a result similar to Theorem 1.2 was proved by G. Conchon-Kerjan in [Con21]. Namely, assuming that \mathcal{G}_n is a *uniformly random* d -regular graph with n vertices, he shows that under the *annealed* probability measure P_{ann} (that is taking into account the randomness of the graph and the field),

$$\lim_{n \rightarrow \infty} P_{\text{ann}} \left(|\mathcal{C}_{\text{max}}^{\mathcal{G}_n, h}|/n \in (\eta(h) - \delta, \eta(h) + \delta) \text{ and } c^{-1} \log n \leq |\mathcal{C}_{\text{sec}}^{\mathcal{G}_n, h}| \leq c \log n \right) = 1, \quad (1.7)$$

and further gives a rather detailed description of the geometry of $\mathcal{C}_{\text{max}}^{\mathcal{G}_n, h}$. Note that the estimate on $|\mathcal{C}_{\text{sec}}^{\mathcal{G}_n, h}|$ in (1.7) is essentially optimal, and better than in (1.6). The arguments of [Con21] are completely different from the ones used in this paper and rely strongly on the assumption that \mathcal{G}_n is a random regular graph, and thus can be revealed progressively using the usual pairing construction. As discussed above, this assumption is stronger than our Assumption 1.1.

Let us now comment on the proof of Theorem 1.2. To explain its main ideas, it is useful to discuss the sprinkling construction for the Bernoulli percolation from [ABS04] first. This construction relies on the fact that a percolation configuration $(\omega^p(x))_{x \in \mathcal{V}_n} \in \{0, 1\}^{\mathcal{V}_n}$ at level p can be obtained as the maximum of two independent Bernoulli configurations ω^{p_1} and ω^{p_2} with the levels p_1, p_2 satisfying $1 - p = (1 - p_1)(1 - p_2)$. For the techniques of [ABS04] to work, it is very important that (a) ω^{p_2} is independent of ω^{p_1} , (b) ω^{p_2} is a Bernoulli percolation, that is the random variables $(\omega^{p_2}(x))_{x \in \mathcal{V}_n}$ are independent, and (c) that the maximum function is monotonous, in particular $\{x : \omega^p(x) = 1\} \supset \{x : \omega^{p_1}(x) = 1\}$. While (c) is important for the sprinkling not to destroy the mesoscopic components of ω^{p_1} , (a) and (b) play a key role in estimating the probability of a certain bad event which needs to be much smaller than e^{-cN_n/m_n} , with m_n denoting the minimal size of mesoscopic components (cf. proof of Proposition 3.1 in [ABS04]). In [ABS04], the proof of this estimate is just a simple large deviation argument for i.i.d. Bernoulli random variables. Unfortunately, a corresponding estimate is mostly simply not true in the setting of correlated Gaussian fields.

Before describing our approach, let us very quickly mention two natural ideas how to adapt the sprinkling construction of [ABS04] to the zero-average Gaussian free field which, unfortunately, cannot easily be converted into a rigorous proof, mostly due to the lack of independence. The first one is to use the existence of many mesoscopic components at a level $h' \in (h, h_*)$ and prove that by lowering the level from h' to h those components merge. This preserves the monotonicity, that is the point (c) from the last paragraph, but completely destroys the independence (a) and (b), making the above mentioned estimate on the bad event essentially impossible to prove. The second one is to write $\Psi^{\mathcal{G}_n}$ as a linear combination $\sqrt{1 - t^2} \Psi'_{\mathcal{G}_n} + t \Psi''_{\mathcal{G}_n}$ (with a small t) of its independent copies $\Psi'_{\mathcal{G}_n}, \Psi''_{\mathcal{G}_n}$. Here, the monotonicity (c) is lost (but probably could be salvaged by some technical work), (a) is preserved, but the correlations of $\Psi''_{\mathcal{G}_n}$ make the estimate on the bad event fail again. Remark also that the zero average property (1.2) excludes writing $\Psi_{\mathcal{G}_n}$ as a sum $X + Y$ of two non-trivial independent fields X, Y such that $Y \geq 0$, or as $\max(X, Y)$ for X, Y independent; both of these decompositions would be desirable for the monotonicity (c).

In this paper we thus develop a new decomposition of the zero-average Gaussian free field which provides enough independence to be useful in a sprinkling argument and which is of independent interest, see Section 3. It is inspired by a similar decomposition of the (usual) Gaussian free field on \mathbb{Z}^d from [DGRS20]. Using this decomposition we

will write $\Psi_{\mathcal{G}_n}$ as a sum of two independent components $\Psi_{\mathcal{G}_n} = \Psi_{\mathcal{G}_n}^1 + \Psi_{\mathcal{G}_n}^2$, where

$$\Psi_{\mathcal{G}_n}^2(x) := t_n \left(Z_0(x) - N^{-1} \sum_{y \in \mathcal{V}_n} Z_0(y) \right), \quad x \in \mathcal{V}_n, \quad (1.8)$$

for some family $(Z_0(x))_{x \in \mathcal{V}_n}$ of i.i.d. Gaussian random variables. Since the field $\Psi_{\mathcal{G}_n}^2$ is essentially an i.i.d. field, up to the zero-average property, this writing preserves (a) and (b) from the above discussion, but gives up on the monotonicity (c). We will deal with the non-monotonicity issue by taking t_n small and by restricting the connected components of the level set to certain subgraphs of \mathcal{G}_n where Z_0 is not too small. These arguments are relatively straightforward and are given in Sections 8, 9.

The decomposition, however, introduces a new problem: the field $\Psi_{\mathcal{G}_n}^1$ is not longer a zero-average Gaussian free field and we thus do not know that it has many mesoscopic components in the whole supercritical phase $h < h_*$. To show this we will use a perturbative argument. More precisely, we use the fact that $\Psi_{\mathcal{G}_n}^1 = \Psi_{\mathcal{G}_n} - \Psi_{\mathcal{G}_n}^2$ (with $\Psi_{\mathcal{G}_n}$ and $\Psi_{\mathcal{G}_n}^2$ dependent!) and that $\Psi_{\mathcal{G}_n}$ has many mesoscopic components at any level $h' \in (h, h_*)$, by [AČ20b, Theorem 5.1]. We then show that, typically, these components are robust to certain perturbations and are thus not destroyed by subtracting $\Psi_{\mathcal{G}_n}^2$. The construction of robust components is based on multi-type branching process arguments developed in [AČ20b]. Its details are given in Sections 5–7.

Acknowledgements. The author wish to thank P.-F. Rodriguez for useful discussions.

2. PRELIMINARIES

In this section we introduce the notation and recall few useful facts that we use throughout the paper. For an arbitrary locally-finite, simple, non-oriented graph G we denote by $V(G)$ and $E(G)$ the sets of its vertices and edges. For $x, y \in V(G)$, we write $x \sim y$ when $(x, y) \in E(G)$, $d_G(\cdot, \cdot)$ denotes their graph distance, and $\deg_G(x)$ the degree of x in G . For any $U \subset V(G)$, $|U|$ stands for its cardinality, and $\partial_G U := \{y \in V(G) \setminus U : \exists x \in U \text{ s.t. } x \sim y\}$ denotes its (outer vertex) boundary. For any $r \geq 0$ and $x \in V(G)$ we define the ball and sphere of radius r around x to be $B_G(x, r) := \{y \in V(G) : d_G(x, y) \leq r\}$ and $S_G(x, r) := \{y \in V(G) : d_G(x, y) = r\}$.

We write \bar{P}_x^G for the canonical law on $V(G)^{\mathbb{N}}$ of the *lazy* simple random walk $X = (X_k)_{k \geq 0}$ on G starting at $x \in V(G)$, and \bar{E}_x^G for the corresponding expectation. Under \bar{P}_x^G , the transition probabilities of X are given by

$$\bar{P}_x^G(X_{k+1} = z \mid X_k = y) = \begin{cases} \frac{1}{2}, & \text{if } z = y, \\ \frac{1}{2 \deg_G(x)}, & \text{if } (z, y) \in E(G). \end{cases} \quad (2.1)$$

If G is a finite connected graph, we denote the unique invariant distribution of X by

$$\pi_G(x) := \frac{\deg_G(x)}{2|E(G)|}. \quad (2.2)$$

The zero-average Green function \bar{G}_G of X and its density are given by

$$\begin{aligned} \bar{G}_G(x, y) &:= \sum_{k \geq 0} (\bar{P}_x^G(X_k = y) - \pi_G(y)), \quad x, y \in V(G), \\ \bar{g}_G(x, y) &:= (\deg_G(y))^{-1} \bar{G}_G(x, y). \end{aligned} \quad (2.3)$$

It is easy to check from the reversibility of the random walk that $\bar{g}_G(x, y)$ is a symmetric function. Zero-average Gaussian free field on G is a centred Gaussian process $(\Psi_G(x))_{x \in V(G)}$ whose law is determined by its covariance function

$$E(\Psi_G(x) \Psi_G(y)) = C_0 \bar{g}_G(x, y) \quad \text{for all } x, y \in V(G). \quad (2.4)$$

The constant C_0 only influences the scaling of the field and is introduced for convenience. If G is d -regular, as in Assumption 1.1(a), it is customary to take $C_0 = d/2$. With this choice,

$$\bar{g}_G(x, y) = \frac{1}{2}\bar{G}_G(x, y) = G_G(x, y), \quad (2.5)$$

where $G_G(\cdot, \cdot)$ is the zero-average Green function of the usual continuous-time random walk on G (the factor $\frac{1}{2}$ disappears due to the laziness), and thus the covariance from (2.4) agrees with the one used in [AČ20b], cf. (2.17) therein.

For any field f on G we denote by $E^{\geq h}(f) := \{x \in V(G) : f(x) \geq h\}$ its level set above the level $h \in \mathbb{R}$.

We use \mathbb{T}_d to denote the d -regular infinite tree with root o . For every $x \in V(\mathbb{T}_d)$ we denote by $\text{desc}(x)$ the set of its direct descendants, and for $x \in V(\mathbb{T}_d) \setminus \{o\}$ we use $\text{anc}(x)$ to denote the direct ancestor of x in \mathbb{T}_d . The Gaussian free field on \mathbb{T}_d is a centred Gaussian process $(\varphi_{\mathbb{T}_d}(x))_{x \in V(\mathbb{T}_d)}$ whose distribution is determined by

$$E(\varphi_{\mathbb{T}_d}(x)\varphi_{\mathbb{T}_d}(y)) = g_{\mathbb{T}_d}(x, y) \quad \text{for all } x, y \in V(\mathbb{T}_d), \quad (2.6)$$

where $g_{\mathbb{T}_d}$ is the Green function of the (usual discrete-time) simple random walk on \mathbb{T}_d .

As mentioned earlier, we consider for fixed $d \geq 3$ the d -regular graphs $(\mathcal{G}_n)_{n \geq 1}$ satisfying Assumption 1.1, and we abbreviate $\mathcal{V}_n = V(\mathcal{G}_n)$, $\mathcal{E}_n = E(\mathcal{G}_n)$. For $r \geq 0$, we say that a vertex $x \in \mathcal{V}_n$ is r -treelike, if there is no cycle in $B_{\mathcal{G}_n}(x, r)$. If x is r -treelike, then we fix a graph isomorphism $\rho_{x,r} : B_{\mathcal{G}_n}(x, r) \rightarrow B_{\mathbb{T}_d}(o, r)$ such that $\rho_{x,r}(x) = o$.

We recall from [AČ20b, Proposition 2.2] that there is $\varepsilon \in (0, 1)$ such that for every $n \geq 1$ and $x, y \in \mathcal{V}_n$,

$$\bar{g}_{\mathcal{G}_n}(x, y) \leq C(d-1)^{-d_{\mathcal{G}_n}(x,y)} + N_n^{-\varepsilon}. \quad (2.7)$$

Finally, Assumption 1.1(a,c) imply (by Cheeger's inequality, for the argument see e.g. [ČTW11, (2.11)]) the uniform isoperimetric inequality for the sequence \mathcal{G}_n :

$$\text{There is } \beta' > 0 \text{ such that } \frac{|\partial \mathcal{G}_n A|}{|A|} \geq \beta' \text{ for all } n \geq 1 \text{ and } A \subset \mathcal{V}_n \text{ with } |A| \leq |\mathcal{V}_n|/2. \quad (2.8)$$

We use c, c', C, \dots to denote positive constants with values changing from place to place and which only depend on the degree d and the constants α and β from Assumption 1.1.

3. DECOMPOSITION OF THE FIELD

The goal of this section is to construct a decomposition of the zero-average Gaussian free field into independent components. We believe that this decomposition is of independent interest. It is the main ingredient of our sprinkling construction, as described in the introduction, but also will be used in Section 4 to construct a new coupling of $\Psi_{\mathcal{G}_n}$ and $\varphi_{\mathbb{T}_d}$. The construction of this decomposition is inspired by a similar decomposition for the usual Gaussian free field on \mathbb{Z}^d from [DGRS20], see Lemma 3.1 therein. However, the zero-average property introduces certain complications making the decomposition less straightforward.

For sake of generality, we consider an arbitrary finite, simple, non-oriented, connected graph $G = (V(G), E(G))$ in this section. That is we do not require that Assumption 1.1 holds.

Recall the definition of Ψ_G from (2.4). To introduce its decomposition we need more notation. We write \tilde{G} for the graph obtained from G by adding an additional vertex to the middle of every edge of G , formally $\tilde{G} = (V(\tilde{G}), E(\tilde{G}))$ with

$$V(\tilde{G}) := V(G) \cup E(G), \quad (3.1)$$

$$E(\tilde{G}) := \{(x, e) : x \in V(G), e \in E(G), e \ni x\}. \quad (3.2)$$

Observe that \tilde{G} is a bipartite graph. For $\tilde{x} \in V(\tilde{G})$, let

$$\tilde{\pi}_G(x) := \deg_{\tilde{G}}(\tilde{x}) = \begin{cases} \deg_G(x), & \text{if } \tilde{x} \in V(G), \\ 2, & \text{if } \tilde{x} \in E(G). \end{cases} \quad (3.3)$$

Let $\tilde{Q}_G(\tilde{x}, \tilde{y}) = \mathbf{1}_{(\tilde{x}, \tilde{y}) \in E(\tilde{G})} / \tilde{\pi}_G(\tilde{x})$, $\tilde{x}, \tilde{y} \in V(\tilde{G})$, be the transition matrix of the *usual* simple random walk on \tilde{G} . \tilde{Q}_G acts on the space $\ell^2(\tilde{\pi}_G)$ by $\tilde{Q}_G f(\tilde{x}) = \sum_{\tilde{y} \in V(\tilde{G})} \tilde{Q}_G(\tilde{x}, \tilde{y}) f(\tilde{y})$. Due to the reversibility, \tilde{Q}_G is a self-adjoint operator on $\ell^2(\tilde{\pi}_G)$. Since \tilde{G} is connected and bipartite, \tilde{Q}_G has simple eigenvalues 1 and -1 with respective eigenfunctions $\mathbf{1}$ and w , where $w(\tilde{x}) = 1$ if $\tilde{x} \in V(G)$ and $w(\tilde{x}) = -1$ if $\tilde{x} \in E(G)$. Denoting by $\|\cdot\|_{\tilde{\pi}_G}$ the $\ell^2(\tilde{\pi}_G)$ -norm and by $\langle \cdot, \cdot \rangle_{\tilde{\pi}_G}$ the corresponding scalar product, we have

$$\|\mathbf{1}\|_{\tilde{\pi}_G}^2 = \|w\|_{\tilde{\pi}_G}^2 = \sum_{\tilde{x} \in V(\tilde{G})} \tilde{\pi}_G(x) = 4|E(G)|, \quad \langle \mathbf{1}, w \rangle_{\tilde{\pi}_G} = 0. \quad (3.4)$$

Let Π_G be the orthogonal projection (in $\ell^2(\tilde{\pi}_G)$) on $\text{span}(\mathbf{1}, w)$. Then $\Pi_G^2 = \Pi_G$ and Π_G is self-adjoint in $\ell^2(\tilde{\pi}_G)$. Moreover, since $\mathbf{1}$ and w are eigenvectors of \tilde{Q}_G , the operators Π_G and \tilde{Q}_G commute, and thus, for every $k \in \mathbb{N}$, the operators $\tilde{Q}_G^k \Pi_G = \Pi_G \tilde{Q}_G^k$ and $(\text{Id} - \Pi_G) \tilde{Q}_G^k$ are self-adjoint as well (here Id stands for the identity operator).

Let $(Z_k(\tilde{x}))_{k \in \mathbb{N}, \tilde{x} \in V(\tilde{G})}$ be independent centred Gaussian random variables with

$$\text{Var } Z_k(\tilde{x}) = C_0 / \tilde{\pi}_G(\tilde{x}), \quad (3.5)$$

defined on a probability space (Ω, \mathcal{A}, P) . Finally, for $x \in V(G)$, set

$$\xi_G^k(x) := \sum_{\tilde{y} \in V(\tilde{G})} ((\text{Id} - \Pi_G) \tilde{Q}_G^k)(x, \tilde{y}) Z_k(\tilde{y}) \stackrel{(\text{not.})}{=} ((\text{Id} - \Pi_G) \tilde{Q}_G^k Z_k)(x), \quad (3.6)$$

$$\tilde{\Psi}_G(x) := \sum_{k \in \mathbb{N}} \xi_G^k(x). \quad (3.7)$$

We now show that (3.7) provides the desired decomposition of Ψ_G .

Proposition 3.1. *The series on the right-hand side of (3.7) converges in $L^2(P)$ and P -a.s., and the law of $\tilde{\Psi}_G$ agrees with the law of Ψ_G , that is $\tilde{\Psi}_G$ is a zero-average Gaussian free field on G .*

Proof. We start by computing the covariances of the fields ξ_G^k . Using the independence of $Z_k(\tilde{x})$'s, the self-adjointness of $(\text{Id} - \Pi_G) \tilde{Q}_G^k$ and the fact that Π_G and \tilde{Q}_G commute, for every $x, y \in V(G)$,

$$\begin{aligned} \text{Cov}(\xi_G^k(x), \xi_G^k(y)) &= \sum_{\tilde{z} \in V(\tilde{G})} ((\text{Id} - \Pi_G) \tilde{Q}_G^k)(x, \tilde{z}) ((\text{Id} - \Pi_G) \tilde{Q}_G^k)(y, \tilde{z}) \frac{C_0}{\tilde{\pi}_G(\tilde{z})} \\ &= \frac{C_0}{\tilde{\pi}_G(y)} \sum_{\tilde{z} \in V(\tilde{G})} ((\text{Id} - \Pi_G) \tilde{Q}_G^k)(x, \tilde{z}) ((\text{Id} - \Pi_G) \tilde{Q}_G^k)(\tilde{z}, y) \\ &= \frac{C_0}{\tilde{\pi}_G(y)} ((\text{Id} - \Pi_G) \tilde{Q}_G^{2k})(x, y). \end{aligned} \quad (3.8)$$

To compute the terms involving Π_G , let $(v_i)_{i=1, \dots, |V(\tilde{G})|}$ be an orthonormal basis of $\ell^2(\tilde{\pi}_G)$ composed by the eigenvectors of \tilde{Q}_G such that $v_1 = \mathbf{1} / \|\mathbf{1}\|_{\tilde{\pi}_G}$ and $v_2 = w / \|w\|_{\tilde{\pi}_G}$, and let $(\lambda_i)_{i=1, \dots, |V(\tilde{G})|}$ be the corresponding eigenvalues. Observe $\Pi_G v_i = 0$ for $i \geq 3$ and that

$$\begin{aligned}
\tilde{Q}_G^{2k}(f) &= \sum_{i=1}^{|V(\tilde{G})|} \langle v_i, f \rangle_{\tilde{\pi}_G} \lambda_i^{2k} v_i. \text{ Hence, for } x, y \in V(G), \\
(\Pi_G \tilde{Q}_G^{2k})(x, y) &= \frac{1}{\tilde{\pi}_G(x)} \langle \mathbf{1}_x, (\Pi_G \tilde{Q}_G^{2k}) \mathbf{1}_y \rangle_{\tilde{\pi}_G} \\
&= \frac{1}{\tilde{\pi}_G(x)} \sum_{i=1}^{|V(\tilde{G})|} \langle \mathbf{1}_x, \langle v_i, \mathbf{1}_y \rangle_{\tilde{\pi}_G} \lambda_i^{2k} \Pi_G v_i \rangle_{\tilde{\pi}_G} = \langle \mathbf{1}_y, v_1 \rangle_{\tilde{\pi}_G} v_1(x) + \langle \mathbf{1}_y, v_2 \rangle_{\tilde{\pi}_G} v_2(x) \quad (3.9) \\
&= \frac{\tilde{\pi}_G(y)}{\|\mathbf{1}\|_{\tilde{\pi}_G}^2} + \frac{\tilde{\pi}_G(y)}{\|w\|_{\tilde{\pi}_G}^2} \stackrel{(3.4)}{=} \frac{2\tilde{\pi}_G(y)}{4|E(G)|} \stackrel{(2.2)}{=} \pi_G(y).
\end{aligned}$$

Hence, by (3.8), (3.9) and (3.3),

$$\text{Cov}(\xi_G^k(x), \xi_G^k(y)) = \frac{C_0}{\deg_G(y)} (\tilde{Q}_G^{2k}(x, y) - \pi_G(y)), \quad x, y \in V(G), k \in \mathbb{N}. \quad (3.10)$$

The matrix \tilde{Q}_G^{2k} restricted to $V(G)$ agrees with the k -step transition matrix of the lazy random walk on G . In particular, due to standard convergence results for Markov chains, $|\tilde{Q}_G^{2k}(x, y) - \pi_G(y)| \leq C e^{-ck}$ for all $x, y \in V(G)$, and thus also $|\text{Cov}(\xi_G^k(x), \xi_G^k(y))| \leq C e^{-ck}$. This implies that the series in (3.7) converges in $L^2(P)$. The a.s. convergence is then standard, e.g. using Kolmogorov's maximal inequality. Finally, (3.10) implies that

$$\begin{aligned}
\text{Cov}(\tilde{\Psi}_G(x), \tilde{\Psi}_G(y)) &= \sum_{k \in \mathbb{N}} \text{Cov}(\xi_G^k(x), \xi_G^k(y)) \\
&= \frac{C_0}{\deg_G(y)} \sum_{k \in \mathbb{N}} (\bar{P}_x(X_k = y) - \pi_G(y)), \quad (3.11)
\end{aligned}$$

which agrees with the covariance of Ψ_G from (2.4). Since $\tilde{\Psi}_G$ is obviously a centred Gaussian field, this completes the proof. \square

4. COUPLING WITH A TREE

We now come back to our original setting of Assumption 1.1 and construct, in Proposition 4.1 below, a coupling between the zero-average Gaussian free field $\Psi_{\mathcal{G}_n}$ and the Gaussian free field $\varphi_{\mathbb{T}_d}$. A similar coupling is provided by Theorem 3.1 of [AČ20b]. However, our Proposition 4.1 has several advantages: First, it has a much simpler proof which is based on the decomposition from Section 3. Second, in its proof we also couple the underlying Z -fields (cf. Remark 4.2 below) which will be important later. And third, in contrast to [AČ20b], we use two independent fields $\varphi_{\mathbb{T}_d}, \varphi'_{\mathbb{T}_d}$ in its statement; this will simplify the application of the coupling in the second moment computation in the proof of Proposition 7.1 below.

For the statement recall from Section 2 that $\rho_{x,r}$ denotes a fixed isomorphism of $B_{\mathcal{G}_n}(x, r)$ and $B_{\mathbb{T}_d}(o, r)$, if $x \in \mathcal{V}_n$ is r -treelike.

Proposition 4.1. *There are $c, C \in (0, \infty)$ such that for all $n, r \in \mathbb{N}$, and for all $x, x' \in \mathcal{V}_n$ which are $2r$ -treelike and satisfy $B_{\mathcal{G}_n}(x, 2r) \cap B_{\mathcal{G}_n}(x', 2r) = \emptyset$ there exists a coupling $\mathbb{Q}_n^{x, x'}$ of $\Psi_{\mathcal{G}_n}$ and two independent Gaussian free fields $\varphi_{\mathbb{T}_d}, \varphi'_{\mathbb{T}_d}$ such that for all $\varepsilon > 0$*

$$\begin{aligned}
&\mathbb{Q}_n^{x, x'} \left[\max \left\{ \max_{y \in B_{\mathcal{G}_n}(x, r)} |\Psi_{\mathcal{G}_n}(y) - \varphi_{\mathbb{T}_d}(\rho_{x, 2r}(y))|, \max_{y \in B_{\mathcal{G}_n}(x', r)} |\Psi_{\mathcal{G}_n}(y) - \varphi'_{\mathbb{T}_d}(\rho_{x', 2r}(y))| \right\} > \varepsilon \right] \\
&\leq C d(d-1)^r \left(\exp \left(-\frac{\varepsilon^2 e^{cr}}{18} \right) + \exp \left(-\frac{\varepsilon^2 N_n}{9(r+1)} \right) \right). \quad (4.1)
\end{aligned}$$

Proof. We use the decomposition of $\Psi_{\mathcal{G}_n}$ from Section 3 and a corresponding decomposition of $\varphi_{\mathbb{T}_d}$. Using the notation of Section 3, let $\tilde{\mathcal{V}}_n := V(\tilde{\mathcal{G}}_n)$, and let $Z = (Z_k(\tilde{x}))_{k \in \mathbb{N}, x \in \tilde{\mathcal{V}}_n}$

be a collection of independent Gaussian random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{Q}_n^{x, x'})$ with variances (cf. (3.5), (3.3), we take $C_0 = d/2$ as explained below (2.4))

$$\text{Var } Z_k(\tilde{x}) = \begin{cases} \frac{1}{2}, & \text{if } \tilde{x} \in \mathcal{V}_n, \\ \frac{d}{4}, & \text{if } \tilde{x} \in \mathcal{E}_n. \end{cases} \quad (4.2)$$

Set $\xi_{\mathcal{G}_n}^k(x) := ((\text{Id} - \Pi_{\mathcal{G}_n})\tilde{Q}_{\mathcal{G}_n}^k Z_k)(x)$ and

$$\Psi_{\mathcal{G}_n}(x) := \sum_{k \geq 0} \xi_{\mathcal{G}_n}^k(x) = \sum_{k \geq 0} ((\text{Id} - \Pi_{\mathcal{G}_n})\tilde{Q}_{\mathcal{G}_n}^k Z_k)(x). \quad (4.3)$$

By Proposition 3.1, $\Psi_{\mathcal{G}_n}$ has the law of zero-average Gaussian free field.

We now introduce a similar decomposition for $\varphi_{\mathbb{T}_d}$, similarly to [DGRS20, Lemma 3.1]. Let $\tilde{\mathbb{T}}_d$ be a graph obtained from \mathbb{T}_d by adding a vertex to the middle of every edge, and let $\mathbf{Z} = (Z_k(\tilde{x}))_{k \in \mathbb{N}, \tilde{x} \in V(\tilde{\mathbb{T}}_d)}$ be a collection of independent Gaussian random variables on the same probability space $(\Omega, \mathcal{A}, \mathbb{Q}_n^{x, x'})$ such that (cf. (4.2))

$$\text{Var } Z_k(\tilde{x}) = \begin{cases} \frac{1}{2}, & \text{if } \tilde{x} \in V(\mathbb{T}_d), \\ \frac{d}{4}, & \text{if } \tilde{x} \in V(\tilde{\mathbb{T}}_d) \setminus V(\mathbb{T}_d). \end{cases} \quad (4.4)$$

Denoting $\tilde{\mathbf{Q}}$ the transition matrix of the usual simple random walk on $\tilde{\mathbb{T}}_d$, we set $\zeta^k(x) := (\tilde{\mathbf{Q}}^k \mathbf{Z}_k)(x)$ and

$$\varphi_{\mathbb{T}_d}(x) := \sum_{k \geq 0} \zeta^k(x) = \sum_{k \geq 0} (\tilde{\mathbf{Q}}^k \mathbf{Z}_k)(x). \quad (4.5)$$

Then $\varphi_{\mathbb{T}_d}$ is a Gaussian free field on \mathbb{T}_d . This can be shown by a straightforward adaptation of the proof for the Gaussian free field on \mathbb{Z}^d from [DGRS20] (or by adapting the proof of Proposition 3.1, leaving out the projection Π_G). By introducing an independent copy $\mathbf{Z}' = (Z'_k(x))_{k \in \mathbb{N}, x \in V(\tilde{\mathbb{T}}_d)}$ of \mathbf{Z} , we further define the field $\varphi'_{\mathbb{T}_d}$ by a formula analogous to (4.5), with \mathbf{Z}' instead of \mathbf{Z} .

Let $\tilde{\rho}_{x, 2r} : B_{\tilde{\mathcal{G}}_n}(x, 4r) \rightarrow B_{\tilde{\mathbb{T}}_d}(o, 4r)$ be the natural extension of the isomorphism $\rho_{x, 2r}$ to the balls in graphs $\tilde{\mathcal{G}}_n$ and $\tilde{\mathbb{T}}_d$. (Note that the ball $B_{\tilde{\mathcal{G}}_n}(x, 4r)$ is related to $B_{\mathcal{G}_n}(x, 2r)$, since in $\tilde{\mathcal{G}}_n$ there are additional vertices in the middle of every edge of \mathcal{G}_n .) We now require that under $\mathbb{Q}_n^{x, x'}$ the underlying fields Z , \mathbf{Z} , and \mathbf{Z}' satisfy the following equalities while otherwise being independent:

$$\begin{aligned} Z_k(\tilde{\rho}_{x, 2r}(\tilde{y})) &= Z_k(\tilde{y}) & \text{for every } k \leq 2r, \tilde{y} \in B_{\tilde{\mathcal{G}}_n}(x, 4r), \\ Z'_k(\tilde{\rho}_{x', 2r}(\tilde{y})) &= Z_k(\tilde{y}) & \text{for every } k \leq 2r, \tilde{y} \in B_{\tilde{\mathcal{G}}_n}(x', 4r). \end{aligned} \quad (4.6)$$

Observe that this can be done without changing the distribution of Z , \mathbf{Z} and \mathbf{Z}' , in particular the assumption $B_{\mathcal{G}_n}(x, 2r) \cap B_{\mathcal{G}_n}(x', 2r) = \emptyset$ is necessary for the independence of \mathbf{Z} and \mathbf{Z}' . The assumption that x is $2r$ -treelike implies that the law of the image by $\tilde{\rho}_{x, 2r}$ of the random walk on $\tilde{\mathcal{G}}_n$ started in $\tilde{y} \in B_{\tilde{\mathcal{G}}_n}(x, 2r)$ and stopped on exiting $B_{\tilde{\mathcal{G}}_n}(x, 4r)$ agrees with the law of the random walk on $\tilde{\mathbb{T}}_d$ started in $\tilde{\rho}_{x, 2r}(\tilde{y})$ and stopped on exiting $B_{\tilde{\mathbb{T}}_d}(o, 4r)$, and that this random walk makes at least $2r$ steps before being stopped. As consequence the corresponding transition probabilities agree in the sense of

$$\tilde{Q}_{\mathcal{G}_n}^k(\tilde{y}, \tilde{y}') = \tilde{\mathbf{Q}}^k(\tilde{\rho}_{x, 2r}(\tilde{y}), \tilde{\rho}_{x, 2r}(\tilde{y}')) \text{ for all } \tilde{y} \in B_{\tilde{\mathcal{G}}_n}(x, 2r), \tilde{y}' \in B_{\tilde{\mathcal{G}}_n}(x, 4r), k \leq 2r. \quad (4.7)$$

From (4.3) and (4.5)–(4.7) it follows that for every $y \in B_{\mathcal{G}_n}(x, r)$

$$\Psi_{\mathcal{G}_n}(y) - \varphi_{\mathbb{T}_d}(\rho_{x, 2r}(y)) = \sum_{k > 2r} \xi_{\mathcal{G}_n}^k(y) + \sum_{0 \leq k \leq 2r} (\Pi_{\mathcal{G}_n} \tilde{Q}_{\mathcal{G}_n}^k Z_k)(y) - \sum_{k > 2r} \zeta^k(\rho_{x, 2r}(y)), \quad (4.8)$$

and a similar equality holds when x is replaced by x' and ζ^k by $\zeta'^k := \tilde{\mathbf{Q}}^k \mathbf{Z}'_k$.

We now estimate the three sums on the right-hand side of (4.8). For the last one, we claim that there is a constant $c > 0$ such that for every $\varepsilon > 0$, $k_0 \geq 1$, and $y \in V(\mathbb{T}_d)$,

$$P\left(\left|\sum_{k>k_0} \zeta^k(y)\right| \geq \frac{\varepsilon}{3}\right) \leq 2 \exp\left(-\frac{\varepsilon^2 e^{ck_0}}{18}\right). \quad (4.9)$$

Indeed, observe that $\zeta^k(y)$, $k \geq 0$, are independent Gaussian random variables with $\text{Var} \zeta^k(y) = \tilde{\mathbf{Q}}^{2k}(y, y)/2$ (which can be proved by a similar computation as in (3.8), recalling $C_0 = d/2$). Therefore,

$$\text{Var}\left(\sum_{k>k_0} \zeta^k(y)\right) = \frac{1}{2} \sum_{k>k_0} \tilde{\mathbf{Q}}^{2k}(y, y) \leq e^{-ck_0}, \quad (4.10)$$

where we used the fact that the lazy random walk $(X_k)_{k \geq 0}$ on \mathbb{T}_d satisfies $\tilde{\mathbf{Q}}^{2k}(y, y) \leq e^{-ck}$, which can easily be proved by observing that $d_{\mathbb{T}_d}(o, X_n)$ is a random walk on \mathbb{N} with a drift pointing away from 0. Claim (4.9) then follows by the usual Gaussian tail estimates.

We proceed similarly for the first sum in (4.8). Using (3.10) and the standard estimate on the convergence to stationarity for finite Markov chains (see e.g. [LPW09, (12.11), p.155]),

$$\text{Var} \xi_{\mathcal{G}_n}^k(x) \stackrel{(3.10)}{=} \frac{1}{2} \left(\tilde{\mathbf{Q}}_{\mathcal{G}_n}^{2k}(x, x) - \frac{1}{N_n} \right) \leq e^{-\lambda_{\mathcal{G}_n} k}, \quad (4.11)$$

where $\lambda_{\mathcal{G}_n} \geq \beta$ is the spectral gap appearing in Assumption 1.1(c) (due to the laziness, there is no 2 in the exponent). Therefore, for every $\varepsilon > 0$, $k_0 \geq 1$, and $y \in \mathcal{V}_n$,

$$P\left(\left|\sum_{k>k_0} \xi_{\mathcal{G}_n}^k(y)\right| \geq \frac{\varepsilon}{3}\right) \leq 2 \exp\left(-\frac{\varepsilon^2 e^{\beta k_0}}{18}\right). \quad (4.12)$$

Finally, for the second sum in (4.8), we claim that for every $\varepsilon > 0$, $k_0 \geq 1$ and $y \in \mathcal{V}_n$,

$$P\left(\left|\sum_{0 \leq k \leq k_0} \sum_{\tilde{z} \in \tilde{\mathcal{V}}_n} (\Pi_{\mathcal{G}_n} \tilde{Q}_{\mathcal{G}_n}^k)(y, \tilde{z}) Z_k(\tilde{z})\right| \geq \frac{\varepsilon}{3}\right) \leq 2 \exp\left(-\frac{\varepsilon^2 N_n}{9(k_0 + 1)}\right). \quad (4.13)$$

Indeed, by the same computation as in (3.8)–(3.9),

$$\text{Var}\left(\sum_{0 \leq k \leq k_0} \sum_{\tilde{z} \in \tilde{\mathcal{V}}_n} (\Pi_{\mathcal{G}_n} \tilde{Q}_{\mathcal{G}_n}^k)(y, \tilde{z}) Z_k(\tilde{z})\right) = \frac{1}{2} (k_0 + 1) \pi_{\mathcal{G}_n}(y) = \frac{k_0 + 1}{2N_n}, \quad (4.14)$$

this follows by the same reasoning as above.

Claim (4.1) then follows from (4.8), (4.9), (4.12), and (4.13) using the triangle inequality, a union bound, and the fact that $|B_{\mathbb{T}_d}(o, r)| = |B_{\mathcal{G}_n}(x, r)| = \frac{d(d-1)^r - 2}{d-2} \leq d(d-1)^r$, if x is $2r$ -treelike. \square

Remark 4.2. Later, it will play the key role that the coupling $\mathbb{Q}_n^{x, x'}$ also couples the underlying Z -fields. In particular, we will use that $\mathbb{Q}_n^{x, x'}$ -a.s.

$$Z_0(y) = Z_0(\rho_{x, 2r}(y)), y \in B_{\mathcal{G}_n}(x, 2r), \quad Z_0(y) = Z_0'(\rho_{x', 2r}(y)), y \in B_{\mathcal{G}_n}(x', 2r). \quad (4.15)$$

which follows directly from (4.6).

5. ROBUST COMPONENTS OF THE GFF ON THE TREE

In what follows, we assume that $\Psi_{\mathcal{G}_n}$, $\varphi_{\mathbb{T}_d}$, and the underlying fields Z , \mathbf{Z} are constructed on some probability space (Ω, \mathcal{A}, P) and (4.3), (4.5) hold. As explained in the introduction (cf. (1.8)), in the sprinkling construction we will write $\Psi_{\mathcal{G}_n}$ as a sum of two independent fields $\Psi_{\mathcal{G}_n}^1$ and $\Psi_{\mathcal{G}_n}^2$. To this end, let $t \in (0, 1)$ be a parameter which will

later depend on n , and write $Z_0 = \sqrt{1-t^2}Z_0^1 + tZ_0^2$, where $Z_0^i = (Z_0^i(x) : x \in \tilde{\mathcal{V}}_n)$, $i \in \{1, 2\}$, are two independent copies of Z_0 . Similarly as above (4.3), we define

$$\xi_{\mathcal{G}_n}^{0,i}(x) := (\text{Id} - \Pi_{\mathcal{G}_n})Z_0^i(x) = Z_0^i(x) - \frac{1}{N_n} \sum_{y \in \mathcal{V}_n} Z_0^i(y), \quad x \in \mathcal{V}_n, i \in \{1, 2\}, \quad (5.1)$$

and set,

$$\Psi_{\mathcal{G}_n}^1(x) := \sqrt{1-t^2}\xi_{\mathcal{G}_n}^{0,1}(x) + \sum_{k \geq 1} \xi_{\mathcal{G}_n}^k(x) \quad \text{and} \quad \Psi_{\mathcal{G}_n}^2(x) := t\xi_{\mathcal{G}_n}^{0,2}(x). \quad (5.2)$$

Then $\Psi_{\mathcal{G}_n} = \Psi_{\mathcal{G}_n}^1 + \Psi_{\mathcal{G}_n}^2$, and $\Psi_{\mathcal{G}_n}^1, \Psi_{\mathcal{G}_n}^2$ are independent.

Further, we introduce two independent copies Z_0^1, Z_0^2 of Z_0 so that $Z_0 = \sqrt{1-t^2}Z_0^1 + tZ_0^2$, and define (cf. (4.5))

$$\varphi_{\mathbb{T}_d}^1(x) := \sqrt{1-t^2}Z_0^1(x) + \sum_{k \geq 1} \zeta^k(x) \quad \text{and} \quad \varphi_{\mathbb{T}_d}^2(x) := tZ_0^2(x). \quad (5.3)$$

Then $\varphi_{\mathbb{T}_d} = \varphi_{\mathbb{T}_d}^1 + \varphi_{\mathbb{T}_d}^2$ and the summands are independent.

The goal of the next three sections is to show that the supercritical level sets of $\varphi_{\mathbb{T}_d}^1$, and as consequence also of $\Psi_{\mathcal{G}_n}^1$, have large connected components. Unfortunately, we cannot apply the results of [AČ20a, AČ20b] directly, because $\Psi_{\mathcal{G}_n}^1$ and $\varphi_{\mathbb{T}_d}^1$ are no longer Gaussian free fields. In this section, we thus show that for any $h < h_\star$ the level set $E^{\geq h}(\varphi_{\mathbb{T}_d})$ of the unmodified field $\varphi_{\mathbb{T}_d}$ has infinite components which are robust to certain perturbations. In the next section, we use this result to show that the level set $E^{\geq h}(\varphi_{\mathbb{T}_d}^1)$ percolates if $h < h_\star$ and t is small enough. Finally, in Section 7, we transfer these result to the field $\Psi_{\mathcal{G}_n}^1$, using the coupling from Section 4.

For the sprinkling construction of Section 9, we need to consider two types of perturbations of $E^{\geq h}(\varphi_{\mathbb{T}_d})$. The first one comes from the field $\varphi_{\mathbb{T}_d}^2$, as already explained, and the second one from an independent Bernoulli percolation. For the latter, let $\iota = (\iota(x))_{x \in V(\mathbb{T}_d)}$ be i.i.d. Bernoulli random variables with $P(\iota(x) = 1) = p$ which are independent of Z and thus of $\varphi_{\mathbb{T}_d}$. The robustness against the perturbation by $\varphi_{\mathbb{T}_d}^2$ involves certain averaging property for $\varphi_{\mathbb{T}_d}$ and is driven by a parameter $\gamma \in [-\infty, 0]$. Formally, for $x \in V(\mathbb{T}_d)$ (recalling that $\text{desc}(x)$ is the set of direct descendants of x in \mathbb{T}_d), let $\mathcal{K}(h, p, \gamma)$ be the set of *robust* vertices in $E^{\geq h}(\varphi_{\mathbb{T}_d})$ defined by

$$\mathcal{K}(h, p, \gamma) := \left\{ x \in \mathcal{V}_n : \varphi_{\mathbb{T}_d}(x) \geq h, \iota(x) = 1, \text{ and } \sum_{y \in \text{desc}(x)} \varphi_{\mathbb{T}_d}(y) \geq \gamma \right\}, \quad (5.4)$$

and let $\mathcal{C}_o^{h,p,\gamma}$ be the connected component of $\mathcal{K}(h, p, \gamma)$ containing the root o . Note that if $p = 1$ and $\gamma = -\infty$, then $\mathcal{C}_o^{h,p,\gamma}$ agrees with the connected component \mathcal{C}_o^h of the level set $E^{\geq h}(\varphi_{\mathbb{T}_d})$ containing the root o . We set

$$\eta(h, p, \gamma) := P(|\mathcal{C}_o^{h,p,\gamma}| = \infty), \quad (5.5)$$

$$\mathcal{S} := \{(h, p, \gamma) \in \mathbb{R} \times [0, 1] \times [-\infty, 0] : \eta(h, p, \gamma) > 0\}. \quad (5.6)$$

The main result of this section is the following proposition which shows that, in the supercritical regime, $\mathcal{C}_o^{h,p,\gamma}$ has similar properties as \mathcal{C}_o^h , cf. [AČ20a, Theorems 5.1, 5.3] or [AČ20b, (2.14), (2.16)].

Proposition 5.1. (a) *If $(h, p, \gamma) \in \mathcal{S}$ and $h' < h, p' > p, \gamma' < \gamma$, then $(h', p', \gamma') \in \mathcal{S}$.*

Moreover, for every $h < h_\star$ there is $p < 1$ and $\gamma > -\infty$ such that $(h, p, \gamma) \in \mathcal{S}$.

(b) *For every (h, p, γ) in the interior \mathcal{S}^0 of \mathcal{S} there is $\lambda_h^{p,\gamma} > 1$ such that*

$$\lim_{k \rightarrow \infty} P(|\mathcal{C}_o^{h,p,\gamma} \cap S_{\mathbb{T}_d}(o, k)| \geq (\lambda_h^{p,\gamma})^k / k^2) = \eta(h, p, \gamma) > 0. \quad (5.7)$$

(c) *The functions $(h, p, \gamma) \mapsto \lambda_h^{p,\gamma}$ and $(h, p, \gamma) \mapsto \eta(h, p, \gamma)$ are continuous on \mathcal{S}^0 .*

Remark 5.2. We expect that \mathcal{S} is open, that is $\mathcal{S}^0 = \mathcal{S}$. Proving this would require to study the critical behaviour of $\mathcal{C}_o^{h,p,\gamma}$ which goes beyond the scope of this paper.

Proof of Proposition 4.1. The proof uses branching process techniques and is a modification of the arguments given in Sections 3–5 of [AČ20a] and in Section 3 of [Szn16]. Here, we only explain how these arguments should be adapted to our setting and leave out the parts that are relatively standard in the context of the (multi-type) branching processes.

We first recall the recursive construction of $\varphi_{\mathbb{T}_d}$ from [AČ20a, Section 2.1]. Define random variables

$$Y_o := \varphi_{\mathbb{T}_d}(o) \quad \text{and} \quad Y_x := \varphi_{\mathbb{T}_d}(x) - \frac{1}{d-1} \varphi_{\mathbb{T}_d}(\text{anc}(x)) \quad \text{for } x \in V(\mathbb{T}_d) \setminus \{o\}. \quad (5.8)$$

Then, by the domain Markov property of $\varphi_{\mathbb{T}_d}$, see [AČ20a, (2.6),(2.7)], $(Y_x)_{x \in V(\mathbb{T}_d)}$ are independent random variables such that $Y_o \sim \mathcal{N}(0, \frac{d-1}{d-2})$ and $Y_x \sim \mathcal{N}(0, \frac{d}{d-1})$ for $x \neq o$.

The definition (5.8) can be written as $\varphi_{\mathbb{T}_d}(o) = Y_o$ and

$$\varphi_{\mathbb{T}_d}(x) = \frac{1}{d-1} \varphi_{\mathbb{T}_d}(\text{anc}(x)) + Y_x \quad \text{for } x \in V(\mathbb{T}_d) \setminus \{o\}. \quad (5.9)$$

The field $\varphi_{\mathbb{T}_d}$ is thus determined by Y_x 's, by applying (5.9) recursively. This also implies that $\varphi_{\mathbb{T}_d}$ can be viewed as a multi-type branching process. Indeed, we can view every $x \in S_{\mathbb{T}_d}(o, k)$ as an individual in the k -th generation of the branching process with an attached type $\varphi_{\mathbb{T}_d}(x) \in \mathbb{R}$. (5.9) can be then rephrased as: every individual x has $d-1$ children (d children if $x = o$) whose types, conditionally on $\varphi_{\mathbb{T}_d}(x)$, are chosen independently according to the distribution $\mathcal{N}(\frac{1}{d-1} \varphi_{\mathbb{T}_d}(x), \frac{d}{d-1})$. This point of view can easily be adapted to \mathcal{C}_o^h , namely by considering the same multi-type branching process but instantly killing all individuals whose type does not exceed h . Relying on this point of view, [AČ20a] investigates the properties of \mathcal{C}_o^h using branching process techniques.

We now modify this construction to apply to $\mathcal{C}_o^{h,p,\gamma}$. In addition to instantly killing the individuals whose type does not exceed h , we also kill individuals x for which $\iota(x) = 0$, and we kill all direct descendants of x if $\sum_{y \in \text{desc}(x)} \varphi_{\mathbb{T}_d}(y) < \gamma$. Then the surviving individuals form a component $\bar{\mathcal{C}}_o^{h,p,\gamma}$ which is slightly larger than $\mathcal{C}_o^{h,p,\gamma}$. More precisely, since we only kill the direct descendants of non-robust vertices, and not those vertices themselves,

$$\bar{\mathcal{C}}_o^{h,p,\gamma} = \mathcal{C}_o^{h,p,\gamma} \cup \left\{ x \in \partial_{\mathbb{T}_d} \mathcal{C}_o^{h,p,\gamma} : \varphi_{\mathbb{T}_d}(x) \geq h, \iota(x) = 1, \sum_{y \in \text{desc}(x)} \varphi_{\mathbb{T}_d}(y) \stackrel{!}{<} \gamma \right\}. \quad (5.10)$$

As consequence, $|\bar{\mathcal{C}}_o^{h,p,\gamma}| = \infty$ iff $|\mathcal{C}_o^{h,p,\gamma}| = \infty$, and $|\bar{\mathcal{C}}_o^{h,p,\gamma} \cap S_{\mathbb{T}_d}(o, k)| \geq a$ implies that $|\mathcal{C}_o^{h,p,\gamma} \cap S_{\mathbb{T}_d}(o, k-1)| \geq a/(d-1)$. Hence, it is sufficient to show claims (b,c) for $\bar{\mathcal{C}}_o^{h,p,\gamma}$ instead of $\mathcal{C}_o^{h,p,\gamma}$ (with an additional constant $(d-1)$). The advantage of the former is that it can be interpreted as a multi-type branching process.

The key role in the investigations of [AČ20a] plays certain operator introduced in [Szn16] in order to give a spectral characterisation of the critical value h_* . This operator is defined as follows, cf. [AČ20a, Section 2.2]: Let ν be the centred Gaussian measure of variance $\frac{d-1}{d-2}$. For $h \in \mathbb{R}$, set

$$(L_h f)(a) := (d-1) \mathbf{1}_{[h, \infty)}(a) E^Y \left[f\left(\frac{a}{d-1} + Y\right) \mathbf{1}_{[h, \infty)}\left(\frac{a}{d-1} + Y\right) \right], \quad f \in L^2(\nu), a \in \mathbb{R}, \quad (5.11)$$

where $Y \sim \mathcal{N}(0, \frac{d}{d-1})$ and E^Y is the expectation with respect to Y . The operator L_h is the ‘mean value’ operator corresponding to \mathcal{C}_o^h when it is viewed as a multi-type branching process, more precisely, for any $x \neq o$ and $a \geq h$,

$$(L_h f)(a) = E \left[\sum_{y \in \mathcal{C}_o^h \cap \text{desc}(x)} f(\varphi_{\mathbb{T}_d}(y)) \mid \varphi_{\mathbb{T}_d}(x) = a, x \in \mathcal{C}_o^h \right]. \quad (5.12)$$

Denoting λ_h the largest eigenvalue of L_h , the critical point h_* is given as the unique solution of the equation $\lambda_h = 1$, see [Szn16, Proposition 3.3]

For $\bar{\mathcal{C}}_o^{h,p,\gamma}$, the corresponding operator has a similar, slightly more complicated, form: For $f \in L^2(\nu)$ and $a \in \mathbb{R}$ (and for $x \neq o$, $a \geq h$ in the formula on the right-hand side of the first line, which is only included to motivate the definition),

$$\begin{aligned} (L_h^{p,\gamma} f)(a) &= E \left[\sum_{x \in \bar{\mathcal{C}}_o^{h,p,\gamma} \cap \text{desc}(x)} f(\varphi_{\mathbb{T}_d}(y)) \middle| \varphi_{\mathbb{T}_d}(x) = a, x \in \bar{\mathcal{C}}_o^{h,p,\gamma} \right] \\ &:= p \mathbf{1}_{[h,\infty)}(a) E^Y \left[\mathbf{1}_{[\gamma,\infty)} \left(\sum_{i=1}^{d-1} \left(\frac{a}{d-1} + Y_i \right) \right) \sum_{i=1}^{d-1} f \left(\frac{a}{d-1} + Y_i \right) \mathbf{1}_{[h,\infty)} \left(\frac{a}{d-1} + Y_i \right) \right], \\ &= p(d-1) \mathbf{1}_{[h,\infty)}(a) E^Y \left[\mathbf{1}_{[\gamma,\infty)} \left(\sum_{i=1}^{d-1} \left(\frac{a}{d-1} + Y_i \right) \right) f \left(\frac{a}{d-1} + Y_1 \right) \mathbf{1}_{[h,\infty)} \left(\frac{a}{d-1} + Y_1 \right) \right], \end{aligned} \quad (5.13)$$

where $(Y_i)_{i=1,\dots,d-1}$ are i.i.d. $\mathcal{N}(0, \frac{d}{d-1})$ and E^Y is the corresponding expectation. Note that $L_h = L_h^{1,-\infty}$.

Contrary to L_h , the operator $L_h^{p,\gamma}$ is not self-adjoint in $L^2(\nu)$. We thus need an additional argument to show that (cf. [Szn16, Proposition 3.1]):

$$\begin{aligned} \text{The value } \lambda_h^{p,\gamma} &:= \|L_h^{p,\gamma}\|_{L^2(\nu)} = \sup\{\langle g, L_h^{p,\gamma} g \rangle_{L^2(\nu)} : \|g\|_{L^2(\nu)} = 1\} \\ &\text{is a simple eigenvalue of } L_h^{p,\gamma}. \text{ Moreover, there is a unique, non-} \\ &\text{negative eigenfunction } \chi_h^{p,\gamma} \in L^2(\nu) \text{ of } L_h^{p,\gamma} \text{ corresponding to } \lambda_h^{p,\gamma} \text{ with} \\ &\|\chi_h^{p,\gamma}\|_{L^2(\nu)} = 1. \end{aligned} \quad (5.14)$$

To show this, observe first that from (5.11), (5.13) it follows that there exist functions $K_h, K_h^{p,\gamma} : [h, \infty)^2 \rightarrow (0, \infty)$ such that, for $a \geq h$ (which is the relevant range since $L_h f(a) = L_h^{p,\gamma} f(a) = 0$ for $a < h$),

$$(L_h f)(a) = \int_{[h,\infty)} K_h(a, y) f(y) \nu(dy), \quad (L_h^{p,\gamma} f)(a) = \int_{[h,\infty)} K_h^{p,\gamma}(a, y) f(y) \nu(dy). \quad (5.15)$$

Moreover, $K_h^{p,\gamma} \leq K_h$ for all admissible values of h, p , and γ . Since L_h is a Hilbert-Schmidt operator on $L^2(\nu)$ (see [Szn16, (3.16)]), it follows that $L_h^{p,\gamma}$ is a Hilbert-Schmidt and thus compact operator on $L^2(\nu)$ as well. By Riesz-Schauder theorem (see e.g. [RS80, Theorem 6.15]), every $\lambda \neq 0$ in the spectrum $\sigma(L_h^{p,\gamma})$ of $L_h^{p,\gamma}$ is an eigenvalue of $L_h^{p,\gamma}$ and 0 is the only possible limit point of $\sigma(L_h^{p,\gamma})$. Since $\lambda_h^{p,\gamma} = \|L_h^{p,\gamma}\|_{L^2(\nu)} = \sup\{|\lambda| : \lambda \in \sigma(L_h^{p,\gamma})\}$, it follows that there is $\lambda \in \mathbb{C}$ with $|\lambda| = \lambda_h^{p,\gamma}$ such that $L_h^{p,\gamma} \chi = \lambda \chi$ for some (possibly complex valued) $\chi \in L^2(\nu)$ with $\|\chi\|_{L^2(\nu)} = 1$.

To prove (5.14), it remains to verify that $\lambda > 0$ and $\chi \geq 0$. If χ is not of the form $\chi = \beta g$ for some $\beta \in \mathbb{C}$ and a real-valued non-negative function g , then $\|\chi\|_{L^2(\nu)} = 1$ and $\langle |\chi|, L_h^{p,\gamma} |\chi| \rangle_{L^2(\nu)} > \lambda_h^{p,\gamma}$ which leads to contradiction with the definition of $\lambda_h^{p,\gamma}$ in (5.14). Hence, $\chi = \beta g$. Since the multiplication by scalars preserves eigenfunctions, we can assume that $\beta = 1$, that is $\chi = g$ is non-negative as required. The equality $L_h^{p,\gamma} \chi = \lambda \chi$ then implies that $\lambda > 0$ as well, completing the proof of (5.14).

The continuity of $\lambda_h^{p,h}$ can be shown by the same arguments as in the proof of [Szn16, (3.20)]. Moreover, from (5.13), (5.14) it follows that $\lambda_h^{p,\gamma}$ is strictly decreasing in h and γ , and strictly increasing in p .

The rest of the proof of Proposition 5.1 follows the lines of [AČ20a] with mostly obvious modifications, frequently relying on the fact that $L_h^{p,\gamma}$ is “smaller” than L_h (in the sense explained under (5.15)). In particular, recalling from (5.6) that \mathcal{S} denotes the set of supercritical parameters, it holds that $\mathcal{S} \supset \{(h, p, \gamma) : \lambda_h^{p,\gamma} > 1\}$, and that (b) holds for every $(h, p, \gamma) \in \mathcal{S}^0$. The continuity of η in the claim (c) can be proved using the

same arguments as in Section 5.1 of [AČ20a]. The first part of claim (a) follows by monotonicity. Finally, by (5.11), (5.13), using the definition of $\lambda_h^{p,\gamma}$ from (5.14), using the continuity of $\lambda_h^{p,\gamma}$

$$\lim_{\gamma \rightarrow -\infty} \lambda_h^{p,\gamma} = p \lim_{\gamma \rightarrow -\infty} \lambda_h^{1,\gamma} = p\lambda_h. \quad (5.16)$$

Hence if $h < h_*$ and thus $\lambda_h > 1$, then there exist $p \in (0, 1)$ and $\gamma \in \mathbb{R}$ with $\lambda_h^{p,\gamma} > 1$, proving the second part of (a). \square

6. PERCOLATION FOR THE PRUNED FIELD ON THE TREE

We now consider the *pruned* field $\varphi_{\mathbb{T}_d}^1$ defined in (5.3) (recall that $\varphi_{\mathbb{T}_d}^1$ implicitly depends on the sprinkling strength t) and show that for $h < h_*$ and t small enough its level set $E^{\geq h}(\varphi_{\mathbb{T}_d}^1)$ percolates.

Let $\mathcal{C}_o^{h,p}(t)$ be the connected component of $\{x \in V(\mathbb{T}_d) : \varphi_{\mathbb{T}_d}(x) \geq h, \varphi_{\mathbb{T}_d}^1(x) \geq h, \iota(x) = 1\}$ containing o , and abbreviate $\eta(h, p) := \eta(h, p, -\infty)$.

Proposition 6.1. *For every $\delta \in (0, 1)$, and $h < h_*$, $p \in [0, 1]$ such that $(h, p, -\infty) \in \mathcal{S}^0$,*

$$\lim_{\substack{k \rightarrow \infty \\ t \rightarrow 0}} P(|\mathcal{C}_o^{h,p}(t) \cap S_{\mathbb{T}_d}(o, k)| \geq (p(1 - \delta)\lambda_h)^k) = \eta(h, p). \quad (6.1)$$

(In the limit we allow $k \rightarrow \infty$ and then $t \rightarrow 0$, or $t \rightarrow 0, k \rightarrow \infty$ together.)

Proof. Since $\mathcal{C}_o^{h,p}(t) \subset \mathcal{C}_o^{h,p,-\infty}$, the left-hand side of (6.1) is bounded from above by $\lim_{k \rightarrow \infty} P(|\mathcal{C}_o^{h,p,-\infty} \cap S_{\mathbb{T}_d}(o, k)| \geq 1) = P(|\mathcal{C}_o^{h,p,-\infty}| = \infty) = \eta(h, p)$, by (5.5), yielding the upper bound in (6.1).

For the lower bound, fix $h' \in (h, h_*)$ and $\gamma > -\infty$ such that $(h', p, \gamma) \in \mathcal{S}^0$ and thus $\eta(h', p, \gamma) > 0$, which is possible by the assumptions. By Proposition 5.1(b),

$$\lim_{k \rightarrow \infty} \mathbb{P}(|\mathcal{C}_o^{h',p,\gamma} \cap S_{\mathbb{T}_d}(o, k)| \geq (\lambda_{h'}^{p,\gamma})^k / k^2) = \eta(h', p, \gamma). \quad (6.2)$$

We now show that when $t > 0$ is small, then subtracting $\varphi_{\mathbb{T}_d}^2$ “does not destroy $\mathcal{C}_o^{h',p,\gamma}$ too much”. To this end, recall from (5.3) that $\varphi_{\mathbb{T}_d}^2$ is an i.i.d. field. However, it is not independent of $\varphi_{\mathbb{T}_d}$, so we need to compute its conditional distribution given $\varphi_{\mathbb{T}_d}$.

Lemma 6.2. *Conditionally on $\varphi_{\mathbb{T}_d}$, $\varphi_{\mathbb{T}_d}^2$ is a Gaussian field determined by*

$$E(\varphi_{\mathbb{T}_d}^2(x) \mid \sigma(\varphi_{\mathbb{T}_d})) = \frac{t^2}{2} \left(\varphi_{\mathbb{T}_d}(x) - \frac{1}{d} \sum_{z \sim x} \varphi_{\mathbb{T}_d}(z) \right), \quad (6.3)$$

$$E(\varphi_{\mathbb{T}_d}^2(x)\varphi_{\mathbb{T}_d}^2(y) \mid \sigma(\varphi_{\mathbb{T}_d})) = \frac{t^2}{2} \delta_{x,y} - \frac{t^4}{4} (\delta_{x,y} - \frac{1}{d} \mathbf{1}_{x \sim y}). \quad (6.4)$$

Proof. By (2.6), (4.4), and (5.3), the fields $\varphi_{\mathbb{T}_d}^2$ and $\varphi_{\mathbb{T}_d}$ are centred jointly Gaussian fields satisfying $E(\varphi_{\mathbb{T}_d}^2(x)\varphi_{\mathbb{T}_d}^2(y)) = E(\varphi_{\mathbb{T}_d}^2(x)\varphi_{\mathbb{T}_d}(y)) = t^2 \delta_{x,y}/2$ and $E(\varphi_{\mathbb{T}_d}(x)\varphi_{\mathbb{T}_d}(y)) = g_{\mathbb{T}_d}(x, y)$ for every $x, y \in V(\mathbb{T}_d)$. Denoting Q the transition matrix of the usual random walk on \mathbb{T}_d , we observe that for every $x, y \in V(\mathbb{T}_d)$

$$\begin{aligned} & \text{Cov}(\varphi_{\mathbb{T}_d}^2(x) - \frac{t^2}{2}((\text{Id} - Q)\varphi_{\mathbb{T}_d})(x), \varphi_{\mathbb{T}_d}(y)) \\ &= \frac{t^2}{2} \left(\delta_{x,y} - \sum_{z \in V(\mathbb{T}_d)} (\text{Id} - Q)(x, z) \text{Cov}(\varphi_{\mathbb{T}_d}(z), \varphi_{\mathbb{T}_d}(y)) \right) \\ &\stackrel{(2.6)}{=} \frac{t^2}{2} \left(\delta_{x,y} - \sum_{z \in V(\mathbb{T}_d)} (\text{Id} - Q)(x, z) g_{\mathbb{T}_d}(z, y) \right) = 0, \end{aligned} \quad (6.5)$$

where in the last equality we used the well-known identity $(\text{Id} - Q)g_{\mathbb{T}_d} = \text{Id}$ for the Green function. It follows that the field $\psi := \varphi_{\mathbb{T}_d}^2 - \frac{t^2}{2}(\text{Id} - Q)\varphi_{\mathbb{T}_d}$ is independent of $\sigma(\varphi_{\mathbb{T}_d})$. Hence,

$$E(\varphi_{\mathbb{T}_d}^2 \mid \sigma(\varphi_{\mathbb{T}_d})) = E(\psi + \frac{t^2}{2}(\text{Id} - Q)\varphi_{\mathbb{T}_d} \mid \sigma(\varphi_{\mathbb{T}_d})) = \frac{t^2}{2}(\text{Id} - Q)\varphi_{\mathbb{T}_d}, \quad (6.6)$$

from which (6.3) follows.

The conditional covariance of $\varphi_{\mathbb{T}_d}^2$ agrees with the covariance of ψ (see e.g. [LG16, Corollary 1.10]), which is

$$\begin{aligned} E(\psi(x)\psi(y)) &= E(\varphi_{\mathbb{T}_d}^2(x)\varphi_{\mathbb{T}_d}^2(y)) - \frac{t^2}{2} \sum_{z \in V(\mathbb{T}_d)} (\text{Id} - Q)(y, z) E(\varphi_{\mathbb{T}_d}^2(x)\varphi_{\mathbb{T}_d}(z)) \\ &\quad - \frac{t^2}{2} \sum_{z \in V(\mathbb{T}_d)} (\text{Id} - Q)(x, z) E(\varphi_{\mathbb{T}_d}^2(y)\varphi_{\mathbb{T}_d}(z)) \\ &\quad + \frac{t^4}{4} \sum_{z, z' \in V(\mathbb{T}_d)} (\text{Id} - Q)(x, z)(\text{Id} - Q)(y, z') E(\varphi_{\mathbb{T}_d}(z)\varphi_{\mathbb{T}_d}(z')). \end{aligned} \quad (6.7)$$

(6.4) then follows by substituting the values of the expectations and by applying once more the above identity for the Green function. \square

We continue with the proof of Proposition 6.1. If $x \in \mathcal{C}_o^{h', p, \gamma} \setminus \{o\}$, then $\varphi_{\mathbb{T}_d}(x) \geq h'$ and $\varphi_{\mathbb{T}_d}(\text{anc}(x)) \geq h'$. Therefore, by the robustness condition (5.4), for $x \in \mathcal{C}_o^{h', p, \gamma} \setminus \{o\}$,

$$\begin{aligned} E(\varphi_{\mathbb{T}_d}^1(x) \mid \sigma(\varphi_{\mathbb{T}_d})) &= E(\varphi_{\mathbb{T}_d}(x) - \varphi_{\mathbb{T}_d}^2(x) \mid \sigma(\varphi_{\mathbb{T}_d})) \\ &\stackrel{(6.3)}{=} \varphi_{\mathbb{T}_d}(x) - \frac{t^2}{2} \left(\varphi_{\mathbb{T}_d}(x) - \frac{1}{d} \sum_{z \sim x} \varphi_{\mathbb{T}_d}(z) \right) \\ &= (1 - \frac{t^2}{2})\varphi_{\mathbb{T}_d}(x) + \frac{t^2}{2d}\varphi_{\mathbb{T}_d}(\text{anc}(x)) + \frac{t^2}{2d} \sum_{z \in \text{desc}(x)} \varphi_{\mathbb{T}_d}(z) \\ &\stackrel{(5.4)}{\geq} h' + t^2(\frac{\gamma}{2d} - \frac{1}{2} + \frac{h'}{2d}). \end{aligned} \quad (6.8)$$

By (6.4), $\text{Var}(\varphi_{\mathbb{T}_d}^1(x) \mid \sigma(\varphi_{\mathbb{T}_d})) \leq ct^2$. For t small, $h' + t(\frac{\gamma}{2d} - \frac{1}{2} - \frac{h'}{2d}) > h$ and thus

$$\lim_{t \downarrow 0} P(\varphi_{\mathbb{T}_d}^1(x) \geq h \mid \sigma(\varphi_{\mathbb{T}_d})) = 1, \quad \text{uniformly for } x \in \mathcal{C}_o^{h', p, \gamma} \setminus \{o\}. \quad (6.9)$$

A similar computation implies that (6.9) holds for $x = o$ as well. In addition, by (6.4), conditionally on $\varphi_{\mathbb{T}_d}$, the random variables $\varphi_{\mathbb{T}_d}^1(x)$, $\varphi_{\mathbb{T}_d}^1(y)$ are independent if $d_{\mathbb{T}_d}(x, y) \geq 2$. By the domination argument of [LSS97], the family $(\mathbf{1}_{[h, \infty)}(\varphi_{\mathbb{T}_d}^1(x)) : x \in \mathcal{C}_o^{h', p, \gamma})$ dominates (conditionally on $\varphi_{\mathbb{T}_d}$) an independent Bernoulli percolation on $\mathcal{C}_o^{h', p, \gamma}$ with parameter $g(t)$ and $g(t) \uparrow 1$ as $t \downarrow 0$. As consequence, $\mathcal{C}_0^{h, p}(t)$ dominates $\mathcal{C}_0^{h', pg(t), \gamma}$, and thus by Proposition 5.1(b),

$$\liminf_{k \rightarrow \infty} P(|\mathcal{C}_o^{h, p}(t) \cap S_{\mathbb{T}_d}(o, k)| \geq (\lambda_{h'}^{pg(t), \gamma})^k / k^2) \geq \eta(h', pg(t), \gamma). \quad (6.10)$$

By (5.16) and Proposition 5.1(c), if $h' > h$ is close to h and t, γ are small enough, then $\lambda_{h'}^{pg(t), \gamma} \geq p(1 - \frac{\delta}{2})\lambda_h$, and since $h' > h$ and $\gamma > -\infty$ are arbitrary, the lower bound for (6.1) follows using the continuity of η from Proposition 5.1(c). \square

7. MANY MESOSCOPIC COMPONENTS FOR THE PRUNED FIELD ON FINITE GRAPHS

As a corollary of Proposition 6.1 and the coupling stated in Proposition 4.1, we now prove the existence of many mesoscopic components for the level set of the field $\Psi_{\mathcal{G}_n}^1$.

To state this result precisely, we need to introduce an additional notation. Let $\bar{Z}_0^2 = (\bar{Z}_0^2(x), x \in \mathcal{V}_n)$ be a copy of Z_0^2 which is independent of Z , Z_0^1 and Z_0^2 , and set (cf. (5.2))

$$\bar{\Psi}_{\mathcal{G}_n}^2(x) := t \left(\bar{Z}_0^2(x) - \frac{1}{N_n} \sum_{y \in \mathcal{G}_n} \bar{Z}_0^2(y) \right). \quad (7.1)$$

The field $\bar{\Psi}_{\mathcal{G}_n}^2$ has the same law as $\Psi_{\mathcal{G}_n}^2$ and thus $\bar{\Psi}_{\mathcal{G}_n} := \Psi_{\mathcal{G}_n}^1 + \bar{\Psi}_{\mathcal{G}_n}^2$ has the same law as $\Psi_{\mathcal{G}_n}$, that is it is a zero-average Gaussian free field on \mathcal{G}_n . For $p > 1/2$ we define $L = L(p) < 0$ by

$$P(\bar{Z}_0^2(x) \geq L) = p. \quad (7.2)$$

We set

$$\bar{\mathcal{V}}_n := \{x \in \mathcal{V}_n : \bar{Z}_0^2(x) \geq L\}, \quad (7.3)$$

and use $\bar{\mathcal{G}}_n$ to denote the subgraph of \mathcal{G}_n induced by $\bar{\mathcal{V}}_n$. Finally, for $x \in \bar{\mathcal{V}}_n$, let $\mathcal{C}_x^h(t)$ be the connected component component of the set $E^{\geq h}(\Psi_{\mathcal{G}_n}^1) \cap E^{\geq h}(\Psi_{\mathcal{G}_n})$ in $\bar{\mathcal{G}}_n$.

To understand the reason for this notation, note that eventually, in Section 9, we will use $\bar{\Psi}_{\mathcal{G}_n}$, and not $\Psi_{\mathcal{G}_n}$, to show that the supercritical level set has a giant component. In particular, we will use the field $\bar{\Psi}_{\mathcal{G}_n}^2$ for the sprinkling. At the sites where this field is very small, it can potentially destroy the connected components of the level set. To avoid this, we will restrict to $\bar{\mathcal{G}}_n$ in our sprinkling construction. It is also useful to compare the definition of $\mathcal{C}_x^h(t)$ with the definition of $\mathcal{C}_o^{h,p}(t)$ in Section 6, in particular note that the role of the percolation ι is taken by the subgraph $\bar{\mathcal{G}}_n$.

Proposition 7.1. *Let $h < h_*$ and let p be such that $(h, p, -\infty) \in \mathcal{S}^0$. Then there exists $c_h \in (0, 1)$ such that for any $\delta > 0$ and any sequence $t_n \downarrow 0$,*

$$\lim_{n \rightarrow \infty} P \left(\sum_{x \in \mathcal{V}_n} \mathbf{1}_{\{|\mathcal{C}_x^h(t_n)| \geq N_n^{c_h}\}} \geq (1 - \delta)\eta(h, p)N_n \right) = 1. \quad (7.4)$$

Proof. The proof follows the steps of Section 5 of [AČ20b] and is an application of the second moment method. Some simplifications, compared to [AČ20b], are due to the fact that our Proposition 4.1 uses two independent copies of $\varphi_{\mathbb{T}_d}$, so we do not need to use the decoupling inequalities for $\varphi_{\mathbb{T}_d}$ as in [AČ20b].

Let $r_n = c_1 \log N_n$ with $c_1 > 0$, and set

$$\begin{aligned} W_n &:= \{x \in \mathcal{V}_n : x \text{ is } 2r_n\text{-treelike}\}, \\ \widetilde{W}_n &:= \{(x, x') \in W_n \times W_n : B_{\mathcal{G}_n}(x, 2r_n) \cap B_{\mathcal{G}_n}(x', 2r_n) = \emptyset\}. \end{aligned} \quad (7.5)$$

By [AČ20b, (5.6), (5.7)], it is possible to fix c_1 is small, such that, for some $c > 0$ and for all n large enough,

$$|W_n| \geq N_n(1 - N_n^{-c}) \quad \text{and} \quad |\widetilde{W}_n| \geq N_n^2(1 - N_n^{-c}). \quad (7.6)$$

Let $\tilde{\mathcal{C}}_x^h(t_n) \subset \mathcal{C}_x^h(t_n)$ be the connected component of $\mathcal{C}_x^h(t_n) \cap B(x, 2r_n)$ containing x . Fix $h' \in (h, h_*)$ so that $(h', p, -\infty) \in \mathcal{S}^0$, which is possible since \mathcal{S}^0 is open. Set $\varepsilon = (h' - h)/2$, $c_h = pc_1\lambda_{h'}/2$, and define events

$$\begin{aligned} A_x^{\mathcal{G}_n, h} &:= \{|\tilde{\mathcal{C}}_x^h(t_n) \cap S_{\mathcal{G}_n}(x, r_n)| \geq N_n^{c_h}\}, \quad \text{for } x \in \mathcal{G}_n, \\ A_o^{\mathbb{T}_d, h} &:= \{|\mathcal{C}_o^{h,p}(t_n) \cap S_{\mathbb{T}_d}(o, r_n)| \geq N_n^{c_h}\}. \end{aligned} \quad (7.7)$$

We now show

$$\lim_{n \rightarrow \infty} P \left(\sum_{x \in W_n} \mathbf{1}_{A_x^{\mathcal{G}_n, h}} \geq (1 - \delta)\eta(h, p)N_n \right) = 1, \quad (7.8)$$

from which (7.4) directly follows.

To show (7.8), for every pair $x, x' \in \widetilde{W}_n$, we use the coupling $\mathbb{Q}_n^{x, x'}$ from Proposition 4.1 (with $r = r_n$) to couple $\Psi_{\mathcal{G}_n}$ with two independent copies of $\varphi_{\mathbb{T}_d}$, $\varphi'_{\mathbb{T}_d}$ of the Gaussian free field on \mathbb{T}_d . By Remark 4.2, this coupling also couples the underlying fields Z_0 ,

Z_0 and Z'_0 as in (4.15). In addition, we write $Z_0 = \sqrt{1 - t_n^2} Z_0^1 + t_n Z_0^2$ and assume that $Z_0^2(y) = Z_0^2(\rho_{x,2r_n}(y))$ for every $y \in B_{\mathcal{G}_n}(x, 2r_n)$, where Z_0^1, Z_0^2 are independent copies of Z_0 . We also use analogous statements for $Z', Z_0'^2$ in the ball $B_{\mathcal{G}_n}(x', 2r_n)$. We also couple the site percolation ι (introduced in the paragraph above (5.4)) and its independent copy ι' with the field \bar{Z}_0^2 so that

$$\begin{aligned}\iota(\rho_{x,2r_n}(y)) &= \mathbf{1}_{[L,\infty)}(\bar{Z}_0^2(y)), & y \in B_{\mathcal{G}_n}(x, 2r_n), \\ \iota'(\rho_{x',2r_n}(y)) &= \mathbf{1}_{[L,\infty)}(\bar{Z}_0^2(y)), & y \in B_{\mathcal{G}_n}(x', 2r_n),\end{aligned}\tag{7.9}$$

which is always possible due to the choice (7.2) of L and since $B_{\mathcal{G}_n}(x, 2r_n)$ and $B_{\mathcal{G}_n}(x', 2r_n)$ are disjoint. Note that (7.9) implies $\rho_{x,2r_n}(\bar{\mathcal{V}}_n \cap B(x, 2r_n)) = \{y \in B_{\mathbb{T}_d}(o, 2r_n) : \iota(y) = 1\}$.

When all these coupling equalities hold and when the coupling $\mathbb{Q}_n^{x,x'}$ succeeds, that is the complement of the event on the left-hand side of (4.1) occurs, then it follows from the definitions of the components $\tilde{\mathcal{C}}_x^h(t_n)$ and $\mathcal{C}_o^{h,p}(t_n)$ that $A_o^{\mathbb{T}_d, h-\varepsilon} \supset A_x^{\mathcal{G}_n, h} \supset A_o^{\mathbb{T}_d, h+\varepsilon}$, and similarly for x' , replacing $A_o^{\mathbb{T}_d, h\pm\varepsilon}$ by their independent copies defined in terms of the field $\varphi'_{\mathbb{T}_d}$. Hence, for $x \in W_n$,

$$P(A_x^{\mathcal{G}_n, h}) \geq P(A_o^{\mathbb{T}_d, h+\varepsilon}) - e(n),\tag{7.10}$$

where $e(n)$ is the probability that the coupling fails. By Proposition 4.1, $e(n)$ is bounded by right-hand side of (4.1) with $r = r_n$, in particular $0 \leq e(n) \leq cN_n^{-k}$ for any $k \in \mathbb{N}$.

Applying (7.10) and Proposition 6.1 (with $h + \varepsilon < h'$ instead of h) we obtain that for every $x \in W_n$, $\liminf_{n \rightarrow \infty} P(A_x^{\mathcal{G}_n, h}) \geq \eta(h + \varepsilon, p)$. As consequence, since $\varepsilon > 0$ is arbitrary, using (7.6) and the continuity of η from Proposition 5.1(c),

$$\liminf_{n \rightarrow \infty} \frac{1}{N_n} E\left(\sum_{x \in W_n} \mathbf{1}_{A_x^{\mathcal{G}_n, h}}\right) \geq \eta(h, p).\tag{7.11}$$

We now compute the variance of the sum in the last display. Expanding it, and then using the coupling $\mathbb{Q}_n^{x,x'}$ again,

$$\begin{aligned}\text{Var}\left(\sum_{x \in W_n} \mathbf{1}_{A_x^{\mathcal{G}_n, h}}\right) &= \sum_{x, x' \in W_n} \left(P(A_x^{\mathcal{G}_n, h} \cap A_{x'}^{\mathcal{G}_n, h}) - P(A_x^{\mathcal{G}_n, h})P(A_{x'}^{\mathcal{G}_n, h})\right) \\ &\leq |(W_n \times W_n) \setminus \widetilde{W}_n| + \sum_{(x, x') \in \widetilde{W}_n} \left(P(A_o^{\mathbb{T}_d, h-\varepsilon})^2 - P(A_o^{\mathbb{T}_d, h+\varepsilon})^2\right) + e(n).\end{aligned}\tag{7.12}$$

Taking $n \rightarrow \infty$, applying (7.6) and Proposition 6.1 several times, we obtain

$$\limsup_{n \rightarrow \infty} \frac{1}{N_n^2} \text{Var}\left(\sum_{x \in W_n} \mathbf{1}_{A_x^{\mathcal{G}_n, h}}\right) \leq \eta(h - \varepsilon, p)^2 - \eta(h + \varepsilon)^2.\tag{7.13}$$

The statement follows by letting $h' \downarrow h$, that is $\varepsilon \downarrow 0$, using the continuity of η . \square

We finish this section by a simple lemma which gives a lower bound on the number of vertices that are contained in small components of the (non-pruned) field $\Psi_{\mathcal{G}_n}$. This lower bound will be used to show the upper bound on $|\mathcal{C}_{\max}^{\mathcal{G}_n, h}|$ in the proof of Theorem 1.2. In its statement we use $\mathcal{C}_x^{\mathcal{G}_n, h}$ to denote connected component of $E^{\geq h}(\Psi_{\mathcal{G}_n})$ containing $x \in \mathcal{V}_n$.

Lemma 7.2. *Let $\mathcal{H}_n := \{x \in \mathcal{V}_n : \mathcal{C}_x^{\mathcal{G}_n, h} \subset B_{\mathcal{G}_n}(x, r_n/2)\}$ with $r_n = c_1 \log N_n$ as in the last proof. Then for every $h < h_*$ and $\delta > 0$*

$$\lim_{n \rightarrow \infty} P(|\mathcal{H}_n| > (1 - \eta(h) - \delta)N_n) = 1.\tag{7.14}$$

Proof. The proof is very similar to the previous one. Due to (7.6) it is sufficient to prove the claim for $|\mathcal{H}_n \cap W_n|$ instead of $|\mathcal{H}_n|$. For $x \in W_n$ define the events $A_x^{\mathcal{G}_n, h} := \{\mathcal{C}_x^{\mathcal{G}_n, h} \subset B_{\mathcal{G}_n}(x, r_n/2)\}$, $A_o^{\mathbb{T}_d, h} := \{\mathcal{C}_o^h \subset B_{\mathbb{T}_d}(o, r_n/2)\}$. Using Proposition 4.1, we can couple those

events so that $A_0^{\mathbb{T}_d, h-\varepsilon} \subset A_x^{\mathcal{G}_n, h} \subset A_0^{\mathbb{T}_d, h+\varepsilon}$. By (1.4), $\lim_{n \rightarrow \infty} P(A_x^{\mathbb{T}_d, h}) = 1 - \eta(h)$. Using the same first and second moment method arguments as in the previous proof, the lemma easily follows. \square

8. EXPANSION PROPERTIES OF REDUCED GRAPHS

Before going to the final sprinkling step, we need a little lemma that show that particular subgraphs of \mathcal{G}_n still have good expanding properties. To this end recall from (7.3) the definition of the subgraph $\bar{\mathcal{G}}_n$. For $K \in \mathbb{R}$, let

$$\hat{\mathcal{V}}_n = \{x \in \mathcal{V}_n : \bar{Z}_0^2(x) \geq L, \Psi_{\bar{\mathcal{G}}_n}^1(x) \geq K\} \subset \bar{\mathcal{V}}_n, \quad (8.1)$$

and let $\hat{\mathcal{G}}_n$ be the subgraph of \mathcal{G}_n induced by $\hat{\mathcal{V}}_n$. The additional condition $\Psi_{\bar{\mathcal{G}}_n}^1(x) \geq K$ will later ensure that, in the sprinkling step, the sites in $\hat{\mathcal{V}}_n$ have a reasonable chance to be in the level set $E^{\geq h}(\bar{\Psi}_{\bar{\mathcal{G}}_n})$ of the field $\bar{\Psi}_{\bar{\mathcal{G}}_n}$ (defined under (7.1)). We will always assume that $K \leq h$, so that the mesoscopic connected components $\mathcal{C}_x^h(t_n)$ (in $\bar{\mathcal{G}}_n$ as considered in Proposition 7.1) are also connected components in $\hat{\mathcal{G}}_n$.

We now show that $\hat{\mathcal{G}}_n$ has good expansion properties, at least when we only consider its large subsets. Recall from (2.8) that β' is the lower bound on the isoperimetric constants of \mathcal{G}_n .

Lemma 8.1. *For every $\delta > 0$, there exist $K_0 = K_0(\delta)$ and $L_0 = L_0(\delta)$ such that for every $K < K_0$ and $L < L_0$*

$$P\left(\inf_{A \subset \mathcal{V}(\hat{\mathcal{G}}_n) : \delta N_n \leq |A| \leq N_n/2} \frac{|\partial_{\hat{\mathcal{G}}_n} A|}{|A|} \geq \frac{\beta'}{2}\right) \geq 1 - N^{-\varepsilon} \quad (8.2)$$

with $\varepsilon > 0$ independent of δ .

Proof. We show that for K, L sufficiently negative, $B_1(n) = \{x \in \mathcal{V}_n : \bar{Z}_0^2(x) < L\}$ and $B_2(n) = \{x \in \mathcal{V}_n : \Psi_{\bar{\mathcal{G}}_n}^1(x) < K\}$ satisfy

$$P(|B_1(n)| + |B_2(n)| \leq \beta' \delta N_n / 2) \geq 1 - N^{-\varepsilon}. \quad (8.3)$$

The claim of the lemma then follows from (2.8). Indeed, on the event in (8.3), for A as in (8.2),

$$|\partial_{\hat{\mathcal{G}}_n} A| \geq |\partial_{\mathcal{G}_n} A| - (|B_1(n)| + |B_2(n)|) \geq \beta' |A| - \beta' \delta N_n / 2 \geq \beta' |A| / 2. \quad (8.4)$$

To prove (8.3), observe first that $|B_1(n)|$ is a binomial random variable with parameters N_n and $p = P(\bar{Z}_0^2(x) < L)$. Hence, by taking L_0 depending on δ sufficiently small, we obtain by the standard large deviation estimates that $P(|B_1(n)| \geq \beta' \delta N_n / 4) \leq e^{-cN_n}$ for all $L < L_0$.

For $B_2(n)$, we use the second moment method again. Observe first that by (5.2),

$$\begin{aligned} \text{Cov}(\Psi_{\bar{\mathcal{G}}_n}^1(x), \Psi_{\bar{\mathcal{G}}_n}^1(y)) &= \text{Cov}(\Psi_{\mathcal{G}_n}(x), \Psi_{\mathcal{G}_n}(y)) - \text{Cov}(\Psi_{\bar{\mathcal{G}}_n}^2(x), \Psi_{\bar{\mathcal{G}}_n}^2(y)) \\ &\stackrel{(2.4)}{=} G_{\mathcal{G}_n}(x, y) - t_n^2 \text{Cov}(\xi_0^2(x), \xi_0^2(y)) \stackrel{(3.10)}{=} G_{\mathcal{G}_n}(x, y) - \frac{t_n^2}{2} (\delta_{x,y} + \frac{1}{N_n}). \end{aligned} \quad (8.5)$$

In particular, using the estimate (2.7) on $G_{\mathcal{G}_n}$, since $t_n \rightarrow 0$, $\sigma_x^2 := \text{Var}(\Psi_{\bar{\mathcal{G}}_n}^1(x)) = G_{\mathcal{G}_n}(x, x) + O(t_n^2) \in (c, c')$ for some $0 < c < c' < \infty$, and if $x \neq y$, for some $\varepsilon \in (0, 1)$,

$$\text{Cov}(\Psi_{\bar{\mathcal{G}}_n}^1(x), \Psi_{\bar{\mathcal{G}}_n}^1(y)) \leq C(d-1)^{-d_{\mathcal{G}_n}(x,y)} + N_n^{-\varepsilon}. \quad (8.6)$$

As consequence, we can fix K_0 small enough so that

$$E(|B_2(n)|) = \sum_{x \in \mathcal{V}_n} P(\Psi_{\bar{\mathcal{G}}_n}^1(x) < K) \leq \sum_{x \in \mathcal{V}_n} e^{-K^2/(2c)} \leq \beta' \delta N_n / 8 \quad (8.7)$$

for every $K \leq K_0$. By the normal comparison lemma, see e.g. [LLR83, Theorem 4.2.1], for $x \neq y \in \mathcal{V}_n$, we then obtain

$$\begin{aligned} P(\Psi_{\mathcal{G}_n}^1(x) \leq K, \Psi_{\mathcal{G}_n}^1(y) \leq K) &= P(\Psi_{\mathcal{G}_n}^1(x) \leq K)P(\Psi_{\mathcal{G}_n}^1(y) \leq K) \\ &\leq C(\text{Cov}(\Psi_{\mathcal{G}_n}^1(x)/\sigma_x, \Psi_{\mathcal{G}_n}^1(y)/\sigma_y) \vee 0) \leq C(\text{Cov}(\Psi_{\mathcal{G}_n}^1(x), \Psi_{\mathcal{G}_n}^1(y)) \vee 0). \end{aligned} \quad (8.8)$$

As consequence, using also that the fact that diameter of \mathcal{G}_n is smaller than $C \log N_n$ and that $|S_{\mathcal{G}_n}(x, r)| \leq d(d-1)^{r-1}$, we obtain

$$\begin{aligned} \text{Var}(|B_2(n)|) &= \sum_{x, y \in \mathcal{V}_n} P(\Psi_{\mathcal{G}_n}^1(x) \leq K, \Psi_{\mathcal{G}_n}^1(y) \leq K) - P(\Psi_{\mathcal{G}_n}^1(x) \leq K)P(\Psi_{\mathcal{G}_n}^1(y) \leq K) \\ &\leq N_n + C \sum_{x \in \mathcal{V}_n} \sum_{r=1}^{C \log n} \sum_{y \in \mathcal{V}_n: d_{\mathcal{G}_n}(x, y)=r} (\text{Cov}(\Psi_{\mathcal{G}_n}^1(x), \Psi_{\mathcal{G}_n}^1(y)) \vee 0) \\ &\stackrel{(8.6)}{\leq} N_n + C \sum_{x \in \mathcal{V}_n} \sum_{r=1}^{C \log n} \sum_{y \in \mathcal{V}_n: d_{\mathcal{G}_n}(x, y)=r} ((d-1)^{-r} + N_n^{-\varepsilon}) \leq N_n^{2-\varepsilon}. \end{aligned} \quad (8.9)$$

By Chebyshev inequality, using (8.7), (8.9), $P(|B_2(n)| \geq \beta' \delta N_n / 4) \leq N_n^{-\varepsilon}$.

Combining the conclusions of the last two paragraphs then implies (8.3) and completes the proof. \square

9. SPRINKLING / PROOF OF THEOREM 1.2

With all preparations of the previous sections, the sprinkling construction is relatively straightforward and follows the steps of [ABS04].

We start by showing that the field $\bar{\Psi}_{\mathcal{G}_n}$ defined under (7.1) contains a giant component of size at least $\eta(h)(1-\delta)N_n$, with probability tending to one as $n \rightarrow \infty$. Since $\bar{\Psi}_{\mathcal{G}_n}$ is a zero-average Gaussian free field on \mathcal{G}_n , this will imply the lower bound on $|\mathcal{C}_{\max}^{\mathcal{G}_n, h}|$ for Theorem 1.2.

For $h < h_*$ and $\delta \in (0, 1/8)$ as in the statement of Theorem 1.2, we fix an arbitrary $h' \in (h, h_*)$ and set $\varepsilon := h' - h$. We further fix K, L small and p close to 1 so that L, p are linked by (7.2) and

$$\begin{aligned} K &= h \wedge K_0(\delta\eta(h')/2), & L &< L_0(\delta\eta(h')/2), \\ (h', p, -\infty) &\in \mathcal{S}^0, & \eta(h', p) &> \eta(h')/2, \end{aligned} \quad (9.1)$$

where $K_0(\delta\eta(h')/2), L_0(\delta\eta(h')/2)$ are as in Lemma 8.1, and the last inequality in (9.1) can be satisfied by Proposition 5.1(c). We let $t_n \rightarrow 0$ slowly so that

$$P(\bar{Z}_0^2(x) \geq t_n^{-1}(h+1-K)) \geq N_n^{-c_{h'}\beta'\delta\eta(h', p)/8}, \quad (9.2)$$

where $c_{h'}$ is as in Proposition 7.1, and β' as in Lemma 8.1.

Due to Lemma 8.1, using also (9.1), we know that:

$$\text{For } \mathcal{A}_n^1 := \left\{ \inf_{A \subset V(\hat{\mathcal{G}}_n): \delta\eta(h') \frac{N_n}{2} < |A| < \frac{N_n}{2}} \frac{|\partial_{\hat{\mathcal{G}}_n} A|}{|A|} \geq \frac{\beta'}{2} \right\} \text{ we have } P(\mathcal{A}_n^1) \geq 1 - N_n^{-c}. \quad (9.3)$$

By Gaussian tail estimates, the zero-averaging term in the definition (7.1) of $\bar{\Psi}_{\mathcal{G}_n}^2$ is negligible with high probability:

$$\text{For } \mathcal{A}_n^2 := \left\{ \left| N_n^{-1} \sum_{y \in \mathcal{V}_n} \bar{Z}_0^2(y) \right| \leq \varepsilon \right\} \text{ we have } P(\mathcal{A}_n^2) \geq 1 - e^{-cN_n}. \quad (9.4)$$

Introducing $m_n := N_n^{c_{h'}}$ to denote the minimal size of mesoscopic components and writing $a_k = (1 - k\delta)\eta(h', p)$ for $k \in \{1, 2\}$, Proposition 7.1 implies that:

$$\text{For } \mathcal{A}_n^3 := \left\{ \sum_{x \in \mathcal{V}_n} \mathbf{1}_{\{|\mathcal{C}_x^h(t_n)| \geq m_n\}} \geq a_1 N_n \right\} \text{ we have } \lim_{n \rightarrow \infty} P(\mathcal{A}_n^3) = 1, \quad (9.5)$$

that is $E^{\geq h'}(\Psi_{\mathcal{G}_n}^1) \cap \hat{\mathcal{V}}_n$ has many mesoscopic components, with high probability. Finally, since $\Psi_{\mathcal{G}_n}^1$ is independent of \bar{Z}_0^2 and the graph $\hat{\mathcal{G}}_n$ depends on \bar{Z}_0^2 only via $\mathbf{1}_{[L, \infty)}(\bar{Z}_0^2(x))$, it follows that:

$$\text{Conditionally on } \sigma(\Psi_{\mathcal{G}_n}^1, \hat{\mathcal{G}}_n), \text{ the random variables } (\bar{Z}_0^2(x))_{x \in \hat{\mathcal{V}}_n} \text{ are i.i.d. distributed as } \mathcal{N}(0, 1/2) \text{ random variable conditioned on being larger than } L. \quad (9.6)$$

In particular, since $L < 0$,

$$p_n := P(\bar{Z}_0(x) \geq t_n^{-1}(h + t_n\varepsilon - K) \mid \sigma(\Psi_{\mathcal{G}_n}^1, \hat{\mathcal{G}}_n), \{x \in \hat{\mathcal{V}}_n\}) \stackrel{(9.2)}{\geq} N_n^{-c_{h'}\beta'\delta\eta(h', p)/8}. \quad (9.7)$$

Assume now that $\mathcal{A}_n := \mathcal{A}_n^1 \cap \mathcal{A}_n^2 \cap \mathcal{A}_n^3$ occurs. On \mathcal{A}_n^3 , we can fix a set of at most $a_1 N_n / m_n$ mesoscopic components of $E^{\geq h'}(\Psi_{\mathcal{G}_n}^1) \cap \hat{\mathcal{V}}_n$ that together contain at least $a_1 N_n$ vertices. Any x in those components satisfies $\bar{Z}_0^2(x) \geq L$ (by definition of $\hat{\mathcal{G}}_n$) and thus (on \mathcal{A}_n^2),

$$\bar{\Psi}_{\mathcal{G}_n}(x) = \Psi_{\mathcal{G}_n}^1(x) + t_n \left(\bar{Z}_0^2(x) - N_n^{-1} \sum_{y \in \mathcal{V}_n} \bar{Z}_0^2(y) \right) \geq h' + t_n(L - \varepsilon) \geq h \quad (9.8)$$

for all n large enough. It follows that these fixed mesoscopic components are contained in $E^{\geq h}(\bar{\Psi}_{\mathcal{G}_n})$. If $E^{\geq h}(\bar{\Psi}_{\mathcal{G}_n})$ has no component of size at least $a_2 N_n$, then one can split these fixed components into two groups A, B , each having at least $\delta\eta(h', p)N_n$ vertices, which are not connected within $E^{\geq h}(\bar{\Psi}_{\mathcal{G}_n})$. There are at most $2^{a_1 N_n / m_n}$ ways to split the fixed mesoscopic components into two groups. By (9.1), $\delta\eta(h', p)N_n > \delta\eta(h')N_n/2$. Therefore, on \mathcal{A}_n^1 , we can use Menger's theorem to show that there are at least $\beta'\delta N_n \eta(h', p)/2$ pairwise vertex-disjoint paths from A to B in $\hat{\mathcal{G}}_n$. Since $\hat{\mathcal{G}}_n$ has at most N_n vertices, at last half of those paths are of length at most $4/\beta'\delta\eta(h', p)$ each. For every $x \in \hat{\mathcal{V}}_n$, $\Psi_{\mathcal{G}_n}^1(x) \geq K$. Therefore, if $\bar{Z}_0^2(x) \geq t_n^{-1}(h + t_n\varepsilon - K)$ and \mathcal{A}_n^2 occurs, then

$$\bar{\Psi}_{\mathcal{G}_n}(x) = \Psi_{\mathcal{G}_n}^1(x) + t_n \left(\bar{Z}_0^2(x) - \frac{1}{N_n} \sum_{x \in \mathcal{V}_n} \bar{Z}_0^2(x) \right) \geq h. \quad (9.9)$$

Hence for the groups A and B being disconnected in $\hat{\mathcal{G}}_n \cap E^{\geq h}(\bar{\Psi}_{\mathcal{G}_n})$, there must be at least one vertex with $\bar{Z}_0^2(x) < t_n^{-1}(h + t_n\varepsilon - K)$ on every of these paths. Due to (9.6) and (9.7), this has probability at most

$$(1 - p_n^{4/\beta'\delta\eta(h', p)})^{\beta'\delta N_n \eta(h', p)/4} \leq \exp(-c(\delta, h', p)N_n^{1-c_{h'}/2}). \quad (9.10)$$

It follows that the probability that \mathcal{A}_n occurs and there is no connected component of $E^{\geq h}(\bar{\Psi}_{\mathcal{G}_n})$ of size at least $a_2 N_n$ (that is there is some partition of the fixed mesoscopic components into groups A and B as above that are disconnected from each other in $E^{\geq h}(\bar{\Psi}_{\mathcal{G}_n})$) is at most

$$2^{a_1 N_n / m_n} \exp(-c(\delta, h', p)N_n^{1-c_{h'}/2}) \leq \exp\{-c'N_n^{1-c_{h'}/2}\}, \quad (9.11)$$

which converges to 0 as $n \rightarrow \infty$.

Together with (9.3)–(9.5), this implies that with probability tending to one with n , $E^{\geq h}(\bar{\Psi}_{\mathcal{G}_n})$ has a connected component of size at least $a_2 N_n = (1 - 2\delta)\eta(h', p)N_n$. Taking h' close to h , p close to 1, using the continuity of $\eta(h, p)$ from Proposition 5.1(c), and recalling that $\bar{\Psi}_{\mathcal{G}_n}$ has the same distribution as $\Psi_{\mathcal{G}_n}$ then proves the lower bound on $|\mathcal{C}_{\max}^{\mathcal{G}_n, h}|$ for our main result (1.6) of Theorem 1.2.

The upper bounds on $|\mathcal{C}_{\max}^{\mathcal{G}_n, h}|$ and $|\mathcal{C}_{\sec}^{\mathcal{G}_n, h}|$ in (1.6) then follow easily from Lemma 7.2 and the lower bound on $|\mathcal{C}_{\max}^{\mathcal{G}_n, h}|$. This completes the proof of Theorem 1.2.

Remark 9.1. We conclude this paper with a short discussion of the assumptions of Theorem 1.2. Assumption 1.1(a) is clearly necessary in all our considerations (besides Section 3).

Assumption 1.1(c), that is the assumption on the spectral gap, is only used to imply the uniform isoperimetric inequality (2.8), and also in (4.11). For our results to be true, we only need (2.8) to hold for macroscopic sets (cf. proof of Lemma 8.1). Also, the argument around (4.11) can be easily adapted if $\lambda_{\mathcal{G}_n} \rightarrow 0$ sufficiently slowly.

Assumption 1.1(b) is only used very implicitly in this paper, namely to ensure that a majority of vertices of \mathcal{G}_n are r_n -treelike with $r_n = c_1 \log N_n$, cf. (7.6) which is proved in [AČ20b, (5.6)] using [ČTW11, Lemma 6.1]. In Sections 7–9 of this paper, we even do not need that r_n grows so quickly. $r_n = C \log \log N_n$ for C sufficiently large would be sufficient for our purposes. Hence Assumption 1.1(b) can be replaced by: For some C sufficiently large,

$$|\{x \in \mathcal{V}_n : x \text{ is } (C \log \log N_n)\text{-treelike}\}| \geq N_n(1 - o(1)). \quad (9.12)$$

For the existence of the giant component (not necessary of size $(1 - \delta)\eta(h)N_n$), the factor $(1 - o(1))$ in the last inequality could even be replaced by a $c \in (0, 1)$.

REFERENCES

- [Abä19] Angelo Abächerli, *Local picture and level-set percolation of the Gaussian free field on a large discrete torus*, Stochastic Process. Appl. **129** (2019), no. 9, 3527–3546. [MR 3985572](#)
- [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](#)
- [AČ20a] Angelo Abächerli and Jiří Černý, *Level-set percolation of the Gaussian free field on regular graphs I: Regular trees*, Electron. J. Probab. **25** (2020), Paper No. 65, 24. [MR 4115734](#)
- [AČ20b] Angelo Abächerli and Jiří Černý, *Level-set percolation of the Gaussian free field on regular graphs II: Finite expanders*, Electron. J. Probab. **25** (2020), Paper No. 130, 39.
- [Con21] Guillaume Conchon-Kerjan, *Anatomy of a Gaussian giant: supercritical level-sets of the free field on random regular graphs*, Preprint, available at [arXiv:2102.10975](#), 2021.
- [ČTW11] Jiří Černý, Augusto Teixeira, and David Windisch, *Giant vacant component left by a random walk in a random d -regular graph*, Ann. Inst. Henri Poincaré Probab. Stat. **47** (2011), no. 4, 929–968. [MR 2884219](#)
- [DGRS20] Hugo Duminil-Copin, Subhajit Goswami, Pierre-François Rodriguez, and Franco Severo, *Equality of critical parameters for percolation of gaussian free field level-sets*, Preprint, available at [arXiv:2002.07735](#), 2020.
- [DR15] Alexander Drewitz and Pierre-François Rodriguez, *High-dimensional asymptotics for percolation of Gaussian free field level sets*, Electron. J. Probab. **20** (2015), 1–39. [MR 3339867](#)
- [KLS20] Michael Krivelevich, Eyal Lubetzky, and Benny Sudakov, *Asymptotics in percolation on high-girth expanders*, Random Structures Algorithms **56** (2020), no. 4, 927–947. [MR 4101348](#)
- [LG16] Jean-François Le Gall, *Brownian motion, martingales, and stochastic calculus*, french ed., Graduate Texts in Mathematics, vol. 274, Springer, [Cham], 2016. [MR 3497465](#)
- [LLR83] M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and related properties of random sequences and processes*, Springer Series in Statistics, Springer-Verlag, New York, 1983.
- [LPW09] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, Providence, RI, 2009, With a chapter by James G. Propp and David B. Wilson. [MR 2466937](#)
- [LSS97] T. M. Liggett, R. H. Schonmann, and A. M. Stacey, *Domination by product measures*, Ann. Probab. **25** (1997), no. 1, 71–95. [MR 1428500](#)
- [RS80] Michael Reed and Barry Simon, *Methods of modern mathematical physics. I*, second ed., Academic Press, Inc., New York, 1980. [MR 751959](#)
- [Szn16] Alain-Sol Sznitman, *Coupling and an application to level-set percolation of the Gaussian free field*, Electron. J. Probab. **21** (2016), 1–26. [MR 3492939](#)

JIRÍ ČERNÝ, DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, UNIVERSITY OF BASEL, SPIEGELGASSE 1, 4051 BASEL, SWITZERLAND